Submitted Deliverable - Still under review

Ref. Ares(2021)6700124 - 29/10/2021

EJP SOIL
European Joint Programme

Towards climate-smart sustainable management of agricultural soils

Deliverable 6.1

Report on harmonized procedures for creation of databases and maps

Due date of deliverable: M14

Actual submission date: 31.03.2021

Resubmission date: 29.10.2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

GENERAL DATA

Grant Agreement: 862695
Project acronym: EJP SOIL
Project title: Towards climate-smart sustainable management of agricultural soils
Project website: www.ejpsoil.eu

Start date of the project: February 1st, 2020
Project duration: 60 months
Name of lead contractor: INRAE

Funding source: H2020-SFS-2018-2020 / H2020-SFS-2019-1
Type of action: European Joint Project COFUND

DELIVERABLE NUMBER:	6.1
DELIVERABLE TITLE:	Report on harmonized procedures for creation of databases and maps
DELIVERABLE TYPE:	Report
WORK PACKAGE N:	WP6
WORK PACKAGE TITLE:	Supporting harmonised soil information and reporting
DELIVERABLE LEADER:	F.M. van Egmond, Wageningen Research
EDITED BY:	F.M. van Egmond and M. Fantappiè
DISSEMINATION LEVEL:	PU

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Authors (alphabetical order):

M.C. Andrenelli[14] (chapter 3)

D. Arrouays[1] (chapter 5)

G. Aust[3] (chapter 5)

Z. Bakacsi[12] (chapter 2)

N.H. Batjes[28] (chapter 3)

A. Bispo[1] (chapter 1, 4)

L. Borůvka[6] (chapter 5)

D. Brus[29] (chapter 4)

J.D. Bulens[2] (chapter 3)

C. Calzolari[27] (chapter 3, 5)

F. De Natale[14] (chapter 3)

C. Di Bene[14] (chapter 3)

L. Donovan[31] (chapter 3)

F.M. van Egmond[2] (chapter 3, 4, 5)

M. Fantappiè[14] (chapter 2, 3, 5)

K. Farkas-Iványi[12] (chapter 2)

L. Gardin[27] (chapter 3)

B. Kempen[28] (chapter 4)

M. Knotters[2] (chapter 4)

G. L'Abate[14] (chapter 3)

C. Lattelais[1] (chapter 3)

C. Le Bas[1] (chapter 3)

E. Leitgeb[3] (chapter 5)

S. Madenoglu[25] (chapter 5)

S. Molnár[12] (chapter 2)

V.L. Mulder[29] (chapter 5)

A.B. Møller[7] (chapter 5)

K. Oorts[30] (chapter 3)

B. Parisse[14] (chapter 3)

V. Penížek[6] (chapter 5)

C. Piccini[14] (chapter 5)

L. Poggio[28] (chapter 5)

N. Saby[1] (chapter 4)

R. Skalsky[20] (chapter 5)

L.M. de Sousa[28] (chapter 3)

B. Stenberg[23] (chapter 5)

K. Teuling[2] (chapter 4)

F. Ungaro[27] (chapter 3, 5)

R. Vašát[6] (chapter 4)

S. Verzandvoort[2] (chapter 1)

D. Walvoort[2] (chapter 5)

J. Wetterlind[23] (chapter 5)

R. Yahiaoui[1] (chapter 3)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Reviewers (alphabetical order):

D. Arrouays[1] (chapter 4)

N. Batjes[28] (chapter 1)

A. Bispo[1] (chapter 2)

L. Boulonne[1] (chapter 4)

C. Chenu[1] (chapter 1, 2, 3, 4, 5)

F. De Natale[14] (chapter 4)

F.M. van Egmond[2] (chapter 1, 2)

M. Fantappiè[14] (chapter 1, 4)

C. Garcia[22] (chapter 5)

G. Heuvelink[28] (chapter 4, 5)

A. Jacobs[10] (chapter 1, 2)

M. Knotters[2] (chapter 5)

P. Lagacherie[1] (chapter 4, 5)

J. Mendes de Jesus[28] (chapter 3)

A.B. Møller[7] (chapter 4)

V.L. Mulder[29] (chapter 4)

O. Nikodemus[15] (chapter 4)

L. Pasztor[12] (chapter 5)

A. Richer de Forges[1] (chapter 5)

J.A. Rodriguez[22] (chapter 4, 5)

F. Schneider[10] (chapter 1, 2, 4)

B. Smreczak[18] (chapter 1, 2, 3, 4, 5)

L.M. de Sousa[28] (chapter 4, 5)

M. C. Trasar Cepeda[22] (chapter 3, 4)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Affiliations

No	Abbreviation	Name	(linked) third party to)
1	INRAE	Institut national de la recherche agronomique	
2	WR	Wageningen Research	
3	BIOS	BIOS Science Austria (Association for the Advancement of Life Sciences)	
4	EV-ILVO	Own capital of the institute for agricultural and fisheries research / Eigen Vermogen van het Instituut voor Landbouw en Visserij Onderzoek (EV ILVO)	
6	CZU	Czech University of Life Sciences Prague / Ceska zemedelska univerzita v Praze	
7	AU	Aarhus Universitet / Aarhus University	
10	Thuenen	Johann Heinrich von Thünen Institut Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei	
12	MTA ATK	Magyar Tudományos Akadémia, Agrártudományi Kutatóközpont	
14	CREA	Council for Agricultural Research and Economics / Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria	
15	UL	University of Latvia, Faculty of Geography and Earth Sciences	
18	IUNG	Institute of Soil Science and Plant Cultivation	
20	NPPC	National Agricultural and Food Centre	
22	INIA	Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, O.A., M.P.	
23	SLU	Swedish University of Agricultural Sciences	
25	TAGEM	Ministry of Agriculture and Forestry/General Directorate of Agricultural Research and Policies	
27	CNR	Consiglio Nazionale delle Ricerch	CREA
28	ISRIC	ISRIC – World Soil Information	WR
29	WU	Wageningen Universiteit	WR
30	VPO	Flemish Planning Bureau for the Environment and Spatial Development / Vlaams Planbureau voor Omgeving	EV-ILVO
31	IR	Inetum-Realdolmen	EV-ILVO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Executive Summary

Many research and policy advice depends on the presence and quality of adequate information. For soil related research and policies, this is existing or newly collected data about soil, soil properties, use, functions, quality and threats. Often the quality and the possible extent of the research and policy advice strongly depends on the available data. The proper *collection, organisation, management and analysis of soil data* towards useful soil information is therefore crucial in any project. This includes performing an inventory of existing data and knowledge, organising and annotating this data such that it is findable, accessible, reusable and if possible interoperable for various projects and purposes, choosing the best strategy for new data collection through sampling or other techniques and then analysing the data towards adequate and understandable spatial soil information products such as maps.

There are however many ways to approach each step of this soil data workflow, it requires a range of different expertise and the best options to choose depend on the aim and scale of the project. This report aims to provide a common basis, a synthesis and a reference for available knowledge on the best practices in this soil data workflow at EJP SOIL partners, aimed at soil institutes.

Data sources

Various recent and ongoing EU and global projects and initiatives have addressed one or more of the topics in the soil data workflow and are therefore presented and discussed in the first chapter and for each topic separately. To understand which soil and soil related data is available in Europe a stocktake on soil data sources was performed. The main conclusions of this stocktake are:

1. Basic soil property (such as soil organic carbon, pH, particle size) databases are available in each country, but some use different measurement methods. If not harmonised before mapping soil properties across Europe, this will result in sudden value changes at national or survey borders of (transboundary issues). A third of the EJP SOIL partners also collect spectral data collection on soil properties.
2. In less than half of the countries data on soil threats are available for soil pollution, compaction, water erosion, and organic matter decline.
3. Many data on soil properties are freely available, but their spatio-temporal resolution varies a lot and often uncertainty quantification is missing. The launch of the European Soil Observatory in 2020 can accelerate a comparison between national databases and LUCAS.
4. There are only a few databases available on measures to control soil erosion. Most partners reported limited access to national soil management databases. This is regulated by the national ministries of agriculture.

In addition to the stocktake on data sources the GSP CountrySIS survey was updated for EJP SOIL partner countries. The main conclusions are:

1. Basic soil data are stored in databases of very different formats, with very different data standards. Their accessibility is variable among countries, among different data owners inside the countries, and for different types of soil data. General soil properties and plant nutrients are always recorded. This is not the case for data on soil salinity, pollution, and contamination.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

2. Not all countries that have soil databases, also have a soil information system (SIS) and/or a soil monitoring system. The maintenance of a SIS needs skilled staff, which is not always available. Some countries reported to have a SIS in 2018, which is no more accessible in 2020. There is a general complaint about the lack of skilled staff, lack of financial resources, lack of time.
3. Other complaints are on the lack of communication/coordination between organizations, which makes it difficult to organize and maintain a national soil database, a connected SIS and a soil monitoring system. Another complaint is on the lack of common standards needed to integrate different soil data sources.
4. Some countries reported the lack of specific legislation for the legal implementation of soil surveys and a soil information system, specifically for soil data protection and data ownership.

Data organisation

In response to the results of the GSP CountrySIS survey, a chapter is dedicated to the ongoing initiatives, background, basic principles and choices for setting up a soil information system while using available standards for data storage, exchange and harmonisation of soil data. Following these structures, it will be easier to organise, store, use and exchange soil data for research, policy and other applications. When setting up a SIS, there is not one single best way to do it because every situation has its own requirements and therefore appropriate choices in architecture, data standards etc. Best practices that apply in all cases include a good documentation of metadata, adherence to existing standards such as INSPIRE, OGC, ISO, Dublin Core etc., and making data findable, accessible, interoperable and reusable (FAIR).

Within the EJP SOIL programme, we aim to set up a distributed soil information system that adheres to and uses the INSPIRE Directive specifications for metadata and soil (Annex III). This means that we choose that data remain at and is curated and updated by its owner (institute) and can be exchanged in a common infrastructure using the INSPIRE soil domain model and appropriate technology. This can be independent from the way partners choose to organise their data. The examples and the overview of harmonisation possibilities show that there is still quite some work to be done before harmonised soil data can be exchanged effortlessly by partners and or member states and the EC/ESDAC/EU Soil Observatory following the INSPIRE model. Currently ongoing activities are aimed at resolving as much as possible the present impediments for full and easy implementation of INSPIRE Soil by partner institutions and member states. This is geared towards at some point in time arriving at a full-fledged standardised decentralised soil information system for Europe that allows harmonisation of soil data for many different applications.

Sampling

Often the existing soil data is not sufficient to answer new questions and new soil data needs to be collected. As becomes clear in the chapter on statistical sampling methods, there is not one sampling method that fits all possible aims and campaigns. Depending on the purpose(s) of the sampling campaign (estimating a mean, mapping, monitoring, gap filling/additional sampling into an existing scheme) a choice needs to be made on the most appropriate design. In general, we can conclude that to estimate global quantities, such as means and totals, probability sampling approaches with design-based inference are more suitable than model-based methods. For sampling for mapping, model based

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

designs are considered more appropriate. In the designs for monitoring, not only spatial variation is a factor, but temporal variation must be taken into account as well.

When choosing a design, a general rule of thumb is to keep a sampling design as simple as possible. The primary concern when designing a monitoring scheme should be to develop an adequate design that makes good use of the available resources and not to construct the perfect, optimal design. Practical convenience and simplicity cannot be sacrificed to achieve optimal statistical efficiency. On the other hand, practical convenience and simplicity should not be the reasons for cumbersome and complicated statistical inferences.

When the aim is to combine data from two designs by far the most important aspect is to know which designs have been used including the details of the construction of the design, such as which strata were used for instance. When the design and, for probability sampling, the inclusion probabilities of the sampling units are known, this can be used to obtain an estimate of a mean or total for the area of interest. How to combine national and European monitoring schemes and other aspects of sampling such as metadata storage and sampling protocol will be elaborated more in deliverable D6.3. An overview of soil monitoring networks in Europe has been published along with suggested options for harmonizing these networks.

Mapping

When the research or policy question requires a map of a soil property, function or threat and the input data is collected it is important to choose the most suitable mapping method. Different intended uses for a map and the availability of existing (in situ point or covariate) data will result in different preferred approaches, there is not one best method. At the same time there are a few general best practices that we advise to adhere to and a stepwise procedure is proposed to select a suitable method. This starts with defining the purpose of the map and inventorying the existing data. Thereby using knowledge of soil forming factors and the SCORPAN model to make effective choices and verifying the quality of the input data and eliminating possible errors, i.e. ‘garbage in is garbage out’. During the entire data collection and mapping process a good documentation of methods, metadata, sampling design strategy and protocol, used data, chosen method, resulting uncertainty metrics and maps, validation of the result and a continuous effort to decrease possible sources of uncertainty are very important and result in a better quality map that can be validated using described methods and a repeatable mapping process. Within EJP SOIL we will adhere to the INSPIRE grid specifications for European and national maps and will aim to reduce transboundary inconsistencies and/ or propose possibilities to address these such as (lab) method harmonisation, combining different sample designs and protocols, using GPS etc.

A good soil data workflow is centred around the defined aims and (research or policy) questions. It uses existing knowledge, experience (from projects, initiatives, literature), and data, a good data organisation, a well-chosen sampling strategy, and the most suitable mapping method followed by validation of the result. This allows to adequately address research and policy questions based on relevant and sufficient quality data, thereby enabling reliable information-based decision making.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

List of Tables

List of Figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

List of acronyms and abbreviations

AT	Austria
CAP	Common Agricultural Policy
CEOS	Committee on Earth Observation Satellites
CIRCASA	Coordination of International Research Cooperation on Soil Carbon Sequestration in Agriculture
CLHS	Conditioned Latin hypercube sampling
COBOL	COmmon Business-Oriented Language
CODASYL	Conference on Data Systems Languages
CR	Cadastral Register
CSM	conventional soil mapping
DEM	Digital Elevation Model
DG-ENV	Directorate General on Environment
Digisoil	Integrated system of data collection technologies for mapping soil properties
DIKW	Data Information Knowledge Wisdom
DSM	Digital Soil Mapping
EBONE	European Biodiversity Observation Network
EC	Electrical Conductivity
ECe	electrical conductivity
ECEC	effective cation exchange capacity
EEA	European Environment Agency
Eionet	European Information and Observation Network
EJP	European Joint Programme
EMI	Electromagnetic Induction
ENVASSO	ENVironmental ASsessment of Soil for mOnitoring
ENVRI	Environmental Research Infrastructures
ENVRI-FAIR	environmental research infrastructures for building FAIR services for research, innovation and society
EO	Earth Observation
EOSC	European Open Science Cloud development
ER	Entity-Relationship
ESA	European Space Agency

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

ESBN	European Soil Bureau Network
ESDAC	European Soil Data Centre
ESDB	European Soil Database
ESP	exchangeable sodium percent
ETC	European Topic Centres
EU	European Union
EUROSOLAN	European and Eurasian Soil Laboratory Network
EUSIS	European Soil Information System
EUSO	EU Soil Observatory
EUSO	EU Soil Observatory
FAIR	Findable, Accessible, Interoperable, Reusable
FAO	Food and Agriculture Organisation of the United Nations
FoI	feature of interest
GDPR	General Data Protection Regulation
GEO	Group of Earth Observation
GEWEX	Global Energy and Water Exchanges Project
GIS	Geographic Information System
GLOSIS	Global Soil Information System
GLOSOLAN	Global Soil Laboratory Network
GML	Geographic Markup Language
GSOC	Global Soil Organic Carbon map
GSP	Global Soil Partnership
GTN-H	Global Terrestrial Network on Hydrology
HWSD	Harmonised World Soil Database
HWSD	Harmonized World Soil Database
HYPRES	Database of Hydraulic Properties of European Soils
IACS	Integrated Administration and Control System
I-ADOPT WG	InteroperAble Descriptions of Observable Property Terminology Working Group
ICP	inductively coupled plasma
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IGAD	Agricultural Data Interest Group

IIASA	International Institute for Applied System Analysis
INSPIRE	Infrastructure for Spatial Information in Europe
IoT	Internet of Things
ISA	Interoperability Solutions for European Public Administrations
ISC	International Science Council
ISMN	International Soil Moisture Network
ISO	International Standards Organisation
iSOIL	Interactions between soil related sciences – Linking geophysics, soil science and digital soil mapping
ISQAPER	interactive soil quality assessment
ISRIC	ISRIC-World Soil Information
ISSCAS	Institute of Soil Science- Chinese Academy of Sciences
ITPS	Intergovernmental Technical Panel on Soils
IUSS	International Union of Soil Science
JRC	Joint Research Centre
JSON	JavaScript Object Notation
LANDMARK	Land Management Assessment Research Knowledge Base
LPIS	Land Parcel Identification System
LPIS	Land Parcel Identification System
LUCAS	Land Use/Cover Area frame Survey
MARS	Monitoring Agricultural Resources
MDA	Model-Driven Architecture
MDD	Model-Driven Development
MDE	Model-Driven Engineering
MG	soil management
MIR	mid-infrared
ML	machine learning
NIR	near-infrared
NPP	net primary productivity
NRC	National Reference Centres
O&M	Observations & Measurements
OC	organic carbon

OGC	Open Geospatial Consortium
OMG	Object Management Group
OWL	Web Ontology Language
PI	Prediction Interval
POP	persistent organic pollutants
PS	Proximal Sensing
PSD	Particle size distribution
PTF	pedotransfer functions
PTRDB	Pedotransfer Rules Database
RDA	Research Data Alliance
RDF	Resource Description Framework
RESOLAN	Regional Soil Laboratory Network
RS	Remote Sensing
RSG	reference soil group
SCORPAN	S: soil, properties of the soil at a point; C: climate, climatic properties of the environment at a point; O: organisms, vegetation or fauna or human activity; R: topography, landscape attributes; P: parent material, lithology; A: age, the time factor; N: space, spatial position.
SCS	spatial coverage sampling
SDG	Sustainable Development Goals
SDG	Sustainable Development Goal
SIEUSOIL	SINO EU Soil Observatory for Intelligent Land use management
SK	Slovakia
SKOS	Simple Knowledge Organization System
SOC	soil organic carbon
SoilIE	OGC Soil Data Interoperability Experiment
SoilSTAT	System for monitoring, forecasting and reporting periodically on the status of global soil resources
SOM	soil organic matter
SP	soil properties
SPADE	soil profile analytical database for Europe
SSA	spatial simulated annealing
TGA	thermogravimetric analysis

TWG	Thematic Working Group
UAV	Unmanned Aerial Vehicle
UML	Unified Modelling Language
UN	United Nations
UNCCD	UN Convention to Combat Desertification
UNFCCC	UN Framework Convention on Climate Change
URI	Uniform Resource Identifiers
USDA	United Stated Department of Agriculture
VNIR	visible near infrared
W3C	World Wide Web Consortium
WCS	Web Coverage Service
WDS	World Data System
WFS	Web Feature Service
WMS	Web Mapping Service
WoSIS	World Soil Information Service
WP	Work Package
WRB	World Reference Base for soil
XML	Extensible Markup Language
XRD	X-ray diffraction

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Introduction

Many research and policy advice depend on the presence and quality of adequate information. For soil related research and policies, for instance towards climate change mitigation and adaptation or the Sustainable Development Goals (SDG's), this is existing or newly collected data about soil, soil properties, use, functions, quality and threats. Often the quality and the possible extent of the research and policy advice strongly depends on the available data. The proper *collection, organisation, management and analysis of soil data* towards useful soil information is therefore crucial in any project. This includes performing an inventory of existing data and knowledge, organising and annotating this data such that it is findable, accessible, reusable and if possible interoperable for various projects and purposes, choosing the best strategy for new data collection through sampling or other techniques and then analysing the data towards adequate and understandable spatial soil information products such as maps.

There are however many ways to approach each step of this soil data workflow, it requires a range of different expertise and the best options to choose depend on the aim and scale of the project. Although this range in expertise is present at most EJP SOIL partner institutions, as EJP SOIL Work Package (WP) 6 (on Supporting harmonised soil information and reporting) we have noticed during the first year of the programme that there are vast differences in knowledge levels on the various topics between partners. This report is an attempt to provide a common basis, a synthesis, on which we can build as WP6 and which can possibly help the internal and external EJP SOIL projects in their work, or other European soil information institutions in general for that matter.

This first deliverable by WP6 was a very collaborative process where many writers and reviewers from many partners participated. This allowed to bring together the latest and best knowledge present on each of these topics within EJP SOIL. The aim was not to write an exhaustive overview of all techniques possible for data organisation, sampling and mapping, nor to provide a complete overview of relevant previous and ongoing EU projects, nor definite choices on how this should be done in EJP SOIL. The aim was to make a synthesis report of the status quo, including the results of the 2020 WP6 stocktake on soil and soil related data sources, the possible choices on each of these topics with their pro's and con's and to provide a lot of references for further information for anyone wanting to dive deeper and get started themselves. It therefore includes background information on methods and terminology but can also be used as a reference for current work, metadata and further reading options and even as a manual that can be consulted when facing choices in the soil data workflow. Most chapters end with a recommendation or direction on each topic within EJP SOIL.

The report starts with an overview of recent and current European and global initiatives and projects that have worked on one or more of the topics addressed in this report (chapter 1). This is further elaborated at the beginning of each topical chapter on data organisation, sampling and mapping. The second chapter presents the results of a stocktake study on available data sources on soil properties, soil monitoring, soil indicators and their metadata. This stocktake was performed by WP6 (task 6.2, 6.3) among EJP SOIL partners in 2020. It also includes updated results of the CountrySIS survey of the Global Soil Partnership for European countries through collaboration with the European Soil Partnership. Chapter 3 is dedicated to soil data organisation. This first provides an overview of the principles, background and terminology of data organisation and harmonisation before applying this in the European context with INSPIRE and showcasing possible implementations and the difficulties for France and Flanders in Belgium. This chapter does not include the topic of data policies as this is the topic of deliverable D6.2, due in summer 2021. Chapter 4 provides an overview of the available methods, choices, and their advantages and disadvantages in sampling design strategies. Special attention is given to options for combining data from different sampling campaigns with different designs and collecting new data in a current design. It does not address other aspects of the sampling protocol or harmonising soil data from different designs. This is the topic of deliverable D6.3, due in summer 2021. Chapter 5 guides though the various aspects of conventional and digital soil mapping, starting at input data, methods, sources of uncertainty, transboundary challenges (national or local) and validation of the result. It provides a best practice list for designing your mapping workflow and ends with general recommendations on soil mapping.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

1 Context and rationale

Authors: A. Bispo and S. Verzandvoort

1.1 Aims

In the European Joint Program *Towards climate-smart sustainable management of agricultural soils* (EJP SOIL), Task 6.1 of WP6 aims to develop a preliminary system for information on agricultural soils. The system should enable to streamline data and information flows between EU countries participating in EJP SOIL and the European Soil Data Centre (ESDAC). The system should handle the harmonization, inventory, reporting and accounting of functional properties of agricultural soils, and complement the soil geodatabase of the Land Use/Cover Area frame Survey (LUCAS). It should also enable sharing of data between national holders of data on agricultural soils, while respecting ownership rights in compliance with the INSPIRE regulation.

Another aim of the information system created in EJP SOIL is to share information between project partners for project purposes. The project coordination team encourages partners to share data as freely as possible, following the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) (Wilkinson, 2016) and the agreements on data operations in the EJP SOIL.

Harmonized procedures are needed to create databases and maps underlying or resulting from the system in order to report and assess different policies, strategies and initiatives (chapter 1.2). This report presents guidelines for data collection, harmonization and mapping in relation to agricultural soils, guidelines for ingestion into a common data model, and sharing methodologies compatible with European and global guidelines. It also includes an overview of current available datasets among project partners.

1.2 Inventory of relevant policies and initiatives

The design and implementation of databases and maps on agricultural soils in Europe are relevant for the execution of several strategies and initiatives that were recently launched by the EU under the European Green Deal and the Horizon Europe framework program, and for international agreements and initiatives in which the EU is a party (e.g. UN SDGs, UN FCCC, UN CCD, FAO, 4pmille). Below, elements from these initiatives that relate to soil data, mapping and harmonisation are summarised.

Table 1.1 summarizes soil data and derived indicators that are required by the policy frameworks. It shows that soil information from field to (supra-)national level is needed to support several global, EU and national policies or initiatives (even if most of those indicators are not yet mandatory). Deriving and mapping soil parameters and indicators to report on soil status change at multiple scale levels will require the development of inter-operable databases with GIS functionality (see also 1.2.9 and 1.2.10 for the description of ongoing cooperation addressing data harmonization and interoperability issues).

Table 1.1 Summary of soil data and derived indicators required by selected policy frameworks.

Policy/initiative	Data/indicator needed	Spatial extent
UN Convention to Combat Desertification	Land cover (land cover change) Land productivity (net primary productivity, NPP) Carbon stocks (soil organic carbon, SOC)	Reporting is expected at country level
UN Framework Convention on Climate Change	Land use change Carbon stocks (soil organic carbon, SOC)	Reporting is expected at country level
UN Sustainable Development Goals		Reporting is expected at country level
SDG 2	Fertility for biomass production (possible indicators are: pH, nutrient content, Organic carbon (OC), cycling of nutrients, water content, soil texture, bulk density)	
SDG 3	Presence of hazardous contaminants (e.g. trace elements, persistent organic pollutants (POPs), texture, OC)	
SDG 6	Hydraulic properties (e.g. bulk density, texture, OC)	
SDG 13	Organic carbon content (e.g. OC, bulk density, coarse fragments)	
SDG 15	Land degradation indicators[1] and soil biodiversity indicators	
FAO-ITPS. Protocol for the assessment of Sustainable Soil Management	*Soil productivity (not a soil indicator, based on yield)* Soil organic carbon (%) Soil physical properties (bulk density) Soil biological activity (soil respiration) Additional indicators may be added (e.g. soil nutrients, soil erosion, soil salinity, soil biodiversity, soil salinity, soil pollution)	Reporting is expected mainly at field scale to compare management options
Recommendations of the EU Mission "Soil Health and food"	Presence of soil pollutants, excess nutrients and salts Vegetation cover Soil organic carbon Soil structure including bulk density and the absence of soil sealing and erosion	Reporting is expected at country level

[1] Indicators under target 15.3.1 align with reporting obligations to the UNCCD under the LDN Target Setting Program (https://www.unccd.int/actions/ldn-target-setting-programme).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

	Soil biodiversity Soil nutrients and pH *Landscape heterogeneity (linked to soil biodiversity[2])* *Area of forest and other wooded lands (not a soil indicator)*	
European Green Deal* **From Farm to Fork** **Biodiversity strategy** **European climate law** **Zero Pollution Action Plan (for Air, Water and Soil)**	Soil organic carbon stock (i.e. OC, bulk density) Soil biodiversity indicators Soil organic carbon (e.g. OC, bulk density) Concentration of hazardous contaminants (e.g. trace elements, POPs, texture, OC)	Reporting is expected at country level
New CAP	Soil organic matter in arable land (indicator C41) Soil erosion by water (indicator C42)	Reporting is expected at country level
4p1000 initiative	Soil organic carbon stock (derived from OC, bulk density, coarse fragments, depth of layer(s))	Reporting is expected at different scales (e.g. field, farm, territory, national)

* To monitor progresses of all policies, the implementation the EU SO is needed requiring a common format for data exchange (at least based on INSPIRE instructions) (see also 1.2.4)

1.2.1 Mission area Soil Health and Food

The Mission area Soil Health and Food[3] formulated the following ambitions or points of attention with regard to the monitoring and assessment of soil condition:

- Exploitation of information ('big data') and technologies such as precision farming, artificial intelligence and remote sensing;
- Integration of data from citizen science, crowd sourcing and multimedia[4];
- Agreed thresholds of soil health indicators for soil type, land use and climate zone;
- Measuring and monitoring techniques of soil health indicators, including proximal and remote sensing;

[2] Landscape heterogeneity is linked to soil biodiversity, and the human impact on the diversity of pedolandscape can be detrimental. Terms used for this impact are 'anthrosolization' and 'entisolization' (https://www.isric.org/sites/default/files/Collection%20management%20plan%20ISRIC%20soil%20reference%20collection_10_2017.pdf and Dazzi and Monteleone (2007)).

[3] https://ec.europa.eu/info/horizon-europe/missions-horizon-europe/soil-health-and-food_en

[4] e.g. multimedia sensors in smart farming to optimize the irrigation process (AlZu'bi et al., 2019)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

- Closer integration between existing pan-European monitoring instruments (e.g. the LUCAS Soil Module) and Member State national programmes;
- Understanding and monitoring footprints on soils outside Europe (e.g. footprints on soil carbon, land degradation, pollution and water use) and developing cooperation, regulations, capacity building and investments.

1.2.2 The European Green Deal

Several key policies and actions proposed under the European Green Deal (European Commission, 2019) require data structures and products related to agricultural soils for their realisation. The most relevant that address agricultural soils are briefly discussed below.

The Farm to Fork Strategy[5] (European Commission, 2020a) proposes to convert the existing Farm Accountancy Data Network into the Farm Sustainability Data Network, with a view to contribute to a wide uptake of sustainable farming practices. Environmental indicators on soil are expected to be part of this (e.g. Poppe and Vrolijk, 2016).

The Farm to Fork Strategy also envisages data collection, processing and analysis in the 'common European agriculture data space'. This data space for the agricultural sector is foreseen in the European Strategy for Data[6] and will be funded under the Digital Europe Program[7]. The Farm to Fork Strategy mentions data on primary production, land use, environmental and other data to be stored in this data space. The motivation is to allow precise and tailored application of 'production approaches' at farm level and monitoring performance of the sector.

The EU Biodiversity Strategy for 2030[8] (European Commission, 2020b) mentions several forms of soil degradation as having negative consequences for the condition and diversity of agroecosystems. In order to restore and protect these, the Strategy indicates that progress is needed in the identification of contaminated soils, the definition of good ecological status of soils, and in the monitoring of soil quality. The Strategy sets out that methods, criteria and standards will be developed in 2021 to describe biodiversity (above- and below-ground) in terms of services, values and sustainable use. This will also include environmental footprints of products and organisations. No references were found in the Strategy policy document to specific data infrastructures or procedures that are envisaged for generating and housing information on agroecosystems.

The Zero Pollution Action Plan for Air, Water and Soil[9] (to be adopted in 2021) addresses soil degradation by pollution. It builds on existing evidence from the European Environment Agency, Eurostat and the Joint Research Centre. Soil pollution by plastics is addressed in the European Strategy for Plastics and the Circular Economy Action Plan.

[5] https://ec.europa.eu/food/farm2fork_en

[6] https://ec.europa.eu/digital-single-market/en/european-strategy-data

[7] https://ec.europa.eu/digital-single-market/en/europe-investing-digital-digital-europe-programme

[8] https://ec.europa.eu/environment/strategy/biodiversity-strategy-2030_nl

[9] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12588-EU-Action-Plan-Towards-a-Zero-Pollution-Ambition-for-air-water-and-soil

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

1.2.3 The new Common Agricultural Policy

The important role of agriculture in managing natural resources is acknowledged and included in the EU's Common Agricultural Policy (CAP), the sustainable management of natural resources being one of its three general objectives. The prevention of soil erosion and improvement of soil management, water management (quantity, quality), restoring, preserving and enhancing biodiversity (landscape, habitats) constitute important elements of this general objective of the CAP. A cross-cutting objective is to modernise the agricultural sector by fostering and sharing knowledge, innovation and digitalisation in agriculture and rural areas, and encouraging uptake (Bas-Defossez and Meredith, 2019). A harmonized data infrastructure for agricultural soils could be instrumental to realise these objectives of the new CAP as it will allow to map vulnerable areas, in particular in relation to sensitivity to soil erosion and the loss of soil organic carbon as requested in a recent report (Augier et al., 2020).

The following soil information is included in the indicators that will be used to inform on the performance of the Common Agricultural Policy: soil organic matter in arable land (indicator C41) and soil erosion by water (indicator C42)[10]. EU countries will need to report (i.e. collect, organize and analyse soil data) on both issues at country level.

The European Commission's proposals on the CAP for the period 2021-2027 set higher ambitions for actions in the domains of environment and climate (European Commission, 2020). These are implemented in the Green Architecture, that is intended to give Member States tools to design and fund environmental and climate schemes. The Green Architecture has the following elements (European Commission, 2020d):

- eco-schemes that provide funding and other incentives for climate- and environment-friendly farming practices;
- support for rural development in agri-environment-climate measures and investments;
- a farm advisory system that will employ a range of economic and environmental data to deliver actual technological and scientific information to farmers (e.g. information related to climate change mitigation and adaptation, biodiversity and protection of water[11]).

A common soil data infrastructure distributed to EU Member States could help to design, monitor and operate these elements as one of the fundamental information carriers.

1.2.4 EU Soil Observatory (EUSO)

The EU Soil Observatory (EUSO)[12] was launched in December 2020 as a platform to provide knowledge and data on soils to the European Commission and its services. It will also serve a broader community of users and support research and innovation on soils. The EUSO will incorporate the European Soil Data Centre (ESDAC)[13], which has been the thematic centre for soil-related data in Europe since 2006.

[10] https://agridata.ec.europa.eu/extensions/DataPortal/cmef_indicators.html

[11] https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/income-support/cross-compliance/fas_en

[12] https://ec.europa.eu/jrc/en/eu-soil-observatory

[13] https://esdac.jrc.ec.europa.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

The ESDAC hosts datasets on soil properties, threats and functions, derived from modelling and monitoring.

The EU policies supported by the EUSO include the new CAP, the Zero Pollution and Farm to Fork Strategies, the Biodiversity Strategy 2030 and the Circular Economy Action Plan (Montanarella and Panagos, 2021). The domains of soil information addressed include soil erosion, soil nutrients, soil organic carbon, soil biodiversity, soil sealing and contaminated sites, as well as soil contamination by agrochemicals, pesticides, organic wastes and industrial emissions. The EUSO will also provide data on soil organic carbon stocks and peatland areas to regularly inform the European Climate Law to achieve net zero greenhouse gas emissions by 2050. The EUSO will support Eurostat in reporting indicators for soil-related Sustainable Development Goals (SDGs), in particular for target 15.3 aimed at achieving land degradation neutrality in the EU. It will contribute to the regular reporting and assessments by IPCC and IPBES on climate change and biodiversity.

The EUSO is supported by the EJP SOIL through the elaboration of an agreement for data collection and sharing on agricultural soils between national data holders and ESDAC. Furthermore, the EJP SOIL supports the EUSO by analysing the options for merging the national strategies of sampling for the monitoring of agricultural soils with the Land Use and Coverage Area frame Survey (LUCAS). Finally, the EJP SOIL supports the EUSO through a software framework that will enable to streamline flows of data on agricultural soils from countries to the ESDAC, if agreed by Member States.

The development of the EU Soil Observatory will start in 2021 under supervision of the JRC. The JRC provides recommendations to the EJP SOIL on how to contribute to the development of the EUSO, with regard to the following subjects (Panagos et al., 2020)[14]:

- Technical aspects, architecture and contents of the LUCAS Soil component (national monitoring networks, spectrometric library and soil biodiversity monitoring);
- Indicators for soil biodiversity, diffuse soil contamination, soil compaction, soil salinization and soil fertility;
- INSPIRE (implementation, data exchange, semantics, and grid system);
- Data flows from EJP SOIL to ESDAC;

The technical advice report from JRC on the collaboration with EJP SOIL (Panagos et al., 2020) also contains a listing of metadata of the available datasets on agricultural soils in ESDAC that could be of use to the EJP SOIL.

1.2.5 EU-China cooperation

The SIEUSOIL project[15] (2019-2022) is an EU Horizon 2020 project that aims to develop sustainable and holistic soil management practices based on a harmonised land information system suitable for diverse climatic and operation conditions at different locations in the EU and China. SIEUSOIL will design, implement and test a shared China-EU Web Observatory platform that will provide Open Linked Data to monitor status and threats of soil, and assist in decision making for sustainable support of agro-ecosystem functions, in view of ongoing climate change. The Observatory platform will support the

[14] https://esdac.jrc.ec.europa.eu/node/68798

[15] https://www.sieusoil.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

wise management of soil at field level through customizable modules and will provide showcases of good practices on soil management both for the EU and China. The final aim will be to support sustainable management of soil, increase land productivity sustainably, reduce crop yield variability across time and space, and support the policy formulation process. Innovative practices and tools will be tested in SIEUSOIL and their impact will be assessed for improved soil fertility and land suitability. Results and developments made by the project will benefit the establishment of the EUSO and allow easier linkage to other international developments such as the Global Soil Information System (GLOSIS) under development by the Global Soil Partnership (GSP).

1.2.6 The European Environment Agency and European Information and Observation Network

The European Environment Agency (EEA) and the European Information and Observation Network (Eionet) provide data and information on environment and climate in Europe for tracking progress on sustainability goals and transitions and policy implementation. The actions and policies supported include in particular the 8th Environmental Action Plan, the European Green Deal and international commitments from Europe. The data and information are collected from a network of institutions from EEA member states and other countries. EEA and Eionet have developed a joint strategy for 2021-2030[16]. Data and information on soil are part of the three areas of work: biodiversity and ecosystems, climate change and adaptation, and human health and the environment (EEA, 2020).

1.2.7 UN Sustainable Development Goals

Several societal research challenges, as described in the UN Sustainable Development Goals (SDG)[17] are directly linked[18] to soil quality and soil management (Keesstra et al., 2020):

- SDG 2: End hunger, achieve food security and improved nutrition, and promote sustainable agriculture;
- SDG 3: Ensure healthy lives and promote well-being for all at all ages (non-communicable diseases, mental health and environmental risks);
- SDG 6: Ensure availability and sustainable management of water and sanitation for all;
- SDG 13: Take urgent action to combat climate change and its impacts;
- SDG 15: Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss.

For member states, reporting on these SDGs to identify progresses or gaps will require monitoring of soil status and evolution.

[16] https://www.eea.europa.eu/about-us/eea-eionet-strategy-2021-2030-1/eea-eionet-strategy-2021-2030

[17] https://sdgs.un.org/goals

[18] Note that SGGs 7 (Ensure access to affordable, reliable, sustainable and modern energy for all), 11 (Make cities and human settlements inclusive, safe, resilient and sustainable) and 12 (Ensure sustainable consumption and production patterns) are also indirectly connected to soil

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

1.2.8 4p1000 initiative

Enhancing sequestration of soil organic carbon recently gained interest by becoming part of the global carbon agenda for climate change mitigation and adaptation; change in SOC stocks is an important indicator for achieving SDG15.3, 'land degradation neutrality'. The '4 per mille' initiative[19] was launched at COP21 by UNFCC in Paris in 2015 to show that agriculture, and in particular wise management of degraded agricultural soils, can play a crucial role for food security and in mitigating climate change. The idea is that an aspirational, annual increase of 4‰ of the global soil organic carbon (SOC) stocks in the top 30 to 40 cm of all agricultural land would counterbalance the annual global rise in atmospheric CO_2. Such increases in SOC stocks are mainly possible in areas of degraded agricultural soils, but will require changes in land management (e.g. adoption of agroecological practices), agroforestry development and restoration of degraded land, as well as enabling economic incentives and policy interventions (Batjes, 2019; Rumpel et al., 2019). Crucial in this context will be the implementation of a consistent system to measure, report and verify soil carbon change in soils (Smith et al., 2020).

1.2.9 GSP, Pillar 4 and 5, GloSIS and ESP

Following the increased recognition that soils are critical for food security and ecosystem services, the Food and Agriculture Organization of the United Nations (FAO)[20] was requested to establish and host a Global Soil Partnership (GSP)[21] in December 2012. The aim of the GSP aim is to develop an effective interactive partnership between governments, regional organizations and institutions on soil-related issues. The initiative is voluntary and does not create any legally binding rights or obligations for its partners or for any other entity under domestic or international law.

One of the key objectives of the GSP is to improve the governance and promote sustainable management of soils, which includes the harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources as well as the enhancement of the quantity and quality of soil data and information (data generation and collection, analysis, validation, reporting, monitoring and integration with other disciplines).

Soil data and information are addressed by Pillar 4[22] of the GSP and by the associated International Network of Soil Information Institutes (INSII)[23], in which most of the national soil information institutes are represented. Pillar 4 is building the Global Soil Information System (GLOSIS), a distributed network to connect and create soil information systems worldwide (see Annex 1). The harmonization of standards on soil data exchange, (laboratory) methods and information is addressed by Pillar 5[24] of the GSP (see chapter 3.1) and by the associated Global Soil Laboratory Network (GLOSOLAN)[25]. Activities

[19] https://www.4p1000.org/
[20] http://www.fao.org/home/en/
[21] http://www.fao.org/global-soil-partnership/en/
[22] http://www.fao.org/global-soil-partnership/pillars-action/4-information-data/en/
[23] http://www.fao.org/global-soil-partnership/insii
[24] http://www.fao.org/global-soil-partnership/pillars-action/5-harmonization/en/
[25] http://www.fao.org/global-soil-partnership/glosolan/en/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

in GSP take place on a global level but also on a regional level. In Europe this is the European Soil Partnership, where activities are also organised according to the Pillar structure.

1.2.10 Infrastructure for Spatial Information in Europe (INSPIRE)

Diverse challenges, such as the availability, quality, organisation, consistency, accessibility, and sharing of spatial information, complicate the formulation of consistent policies in the European Union. These challenges are experienced across the various levels of public authority. Hence, it is necessary to ensure better coordination between the users and providers of spatial information. In this respect, Directive 2007/2/EC of the European Parliament and of the Council, adopted on 14 March 2007, is aimed at establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)[26] to support environmental policies and activities that have an impact on the environment.

INSPIRE is based on the infrastructures for spatial information that are created and maintained by the Member States. To support the establishment of a European infrastructure, implementing rules have been specified for the following components: metadata, interoperability of spatial data sets (as described in Annexes I, II, III of the Directive) and spatial data services, network services, data and service sharing, and monitoring and reporting procedures. Soils are part of Annex III[27], and technical guidelines for data specifications on soils were developed by the Thematic Working Group in 2013 (TWG-Soil, 2013). Details on INSPIRE and its implications for the development of the soil information system for EJP SOIL are given in chapter 3.1.

1.3 Managing soil information: experiences from research projects, programs and platforms

Soil is a complex system at the intersection of the atmosphere, lithosphere, hydrosphere and biosphere. It is critical to food production and key to sustainable living environments through its support of important societal and ecosystem services such as water storage and purification, climate change mitigation, biodiversity preservation or human health.

A better management of soils requires knowledge on their distribution, nature and status. Such information is generally obtained by:

- describing, collecting and analysing soils according to (supra)national standards;
- storing the data according to (supra)national standards;
- developing indicators (based on raw data) and interpretation values (e.g. threshold or reference values) for the status of soil quality (or health);
- providing the results in user-friendly formats (e.g. raw data for farmers or researchers, maps for regional planners, statistical indicators for (supra)national reporting).

---

[26] https://inspire.ec.europa.eu/

[27] https://inspire.ec.europa.eu/Themes/127/2892

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Several research projects, programs and platforms have already addressed these requirements, and their recommendations should be considered during the set-up of the soil information system for EJP SOIL (see Annex 1 and chapter 2-5). Some are briefly introduced below.

1.3.1 Available soil data and information systems

Several countries in Europe have soil information systems with soil datasets at national level and corresponding soil monitoring networks. An overview of datasets and systems known to the EJP SOIL is given in chapter 2. At the European or global level, repositories for soil-related data and soil information systems already exist in the form of LUCAS (Orgiazzi et al, 2019)[28], WoSIS (Batjes et al, 2020)[29], SoilGrids (De Sousa et al., 2020)[30] and GloSIS (Yigini and Van Egmond, 2020)[31]. The European Soil Data Centre (ESDAC[32]) developed a portal for soil datasets and derived products for Europe (see also chapter 5). Alternatively, ISRIC maintains an overview of soil geographic databases[33] for the world as well as a soil data hub[34] containing local, regional and global datasets and maps from around the world (see chapter 3.1.1.5).

1.3.2 Procedures for creating soil databases and exchanging soil data

Procedures for exchanging and sharing of (structured) data in the European Union were regulated in the INSPIRE Directive. Several projects developed guidelines for the description and harmonisation[35] of soil-related information, or designs for the construction of open access databases and portals (ENVASSO[36], GS Soil[37], SIEUSOIL[38], CIRCASA[39], ENVRI-FAIR[40]; Annex 1). An ISO standard was also produced (ISO 28258, 2013).

1.3.3 Development of indicators for the assessment of soil quality/health, soil threats, soil functions and related ecosystems services

Several projects have defined specifications for collecting structuring data, and recommended procedures for calculating and mapping indicators (ENVASSO, RECARE, Landmark; Annex 1). These will

---

[28] https://esdac.jrc.ec.europa.eu/content/lucas-topsoil-survey-methodology-data-and-results
[29] https://www.isric.org/explore/wosis
[30] https://www.isric.org/explore/soilgrids, https://soilgrids.org/
[31] http://www.fao.org/global-soil-partnership/areas-of-work/soil-information-and-data/en/
[32] https://esdac.jrc.ec.europa.eu/
[33] https://www.isric.org/explore/soil-geographic-databases
[34] https://data.isric.org/
[35] mechanisms for soil analysis, the description, classification and mapping of soil, interpretation and exchange of digital soil data; source: GSP Pillar 5 plan of action, http://www.fao.org/3/a-az922e.pdf (see chapter 3.4)
[36] https://esdac.jrc.ec.europa.eu/projects/envasso
[37] https://www.eurogeosurveys.org/projects/gssoil/
[38] https://www.sieusoil.eu/newsletter-3/#1596610924468-29c0a5d7-2dc9
[39] https://www.circasa-project.eu/
[40] https://envri.eu/home-envri-fair/

be detailed in deliverable D6.5 of the EJP SOIL as well as within the SIREN project[41] (recently started as an answer to the 1st internal call of EJP SOIL). Note that relevant indicators for the EJP SOIL should be identified at an early stage as they will define which data need to be collected, as well as the spatial and temporal extent.

1.3.4 Sampling for mapping and monitoring of soils

The main framework for monitoring of top soils in the European Union is the Land Use/Cover Area frame statistical Survey Soil (LUCAS Soil)[42]. Its aim is to derive policy relevant statistics on the impact of land management on soil characteristics. LUCAS uses a consistent methodology.

Alternatively, several systems exist for the collection of data on soils, land cover and land use at the regional, national and European level using field observations or in situ data and satellite imagery, such as the Copernicus Land Monitoring Service, the European Biodiversity Observation Network, the ESA-WORLDSOILS system and components of the Monitoring Agricultural Resources facility of DG Agriculture (LPIS, MARS Crop Yield Forecasting). Monitoring schemes and applications for the assessment of soil quality and management were developed specifically for agricultural soils in the EU research projects ENVASSO, LANDMARK and ISQAPER (Annex 1). At national level member states also developed their own monitoring schemes, i.e. often using national standards (see chapter 2).

1.3.5 Procedures for creating maps

Procedures for creating maps of soil types and properties have evolved from conventional landscape-based soil mapping into predictive (digital) soil mapping, as discussed by Ma et al. (2019). For predictive soil mapping at national, continental and global level the following are needed: soil point observations representing different pedo-climatic regions, environmental and remote sensing-based covariates representing the main soil forming factors, and a geo-statistical model calibrated at an appropriate spatial and temporal resolution. Evaluation of maps using statistical approaches as well as expert opinion are needed in both modelling and evaluation phases.

The Global Soil Information System of the GSP (GLOSIS), for example, enables the creation of global maps of soil properties using predictive soil mapping (e.g. the Global Soil Organic Carbon Map[43]). Specifications for the production of grids of soil properties at global level, at 90 m resolution, are being developed by the GlobalSoilMap IUSS Working Group[44] (GlobalSoilMap 2015[45]). Alternatively, SoilGrids is a system aimed at global predictive soil mapping (De Sousa et al., 2020). Other methods to generate maps of soil and terrain attributes from 'best available' soil survey data and remote and proximal sensing were developed during the research projects e-SOTER, iSOIL and Digisoil.

---

[41] https://projects.au.dk/research-projects/siren/
[42] https://esdac.jrc.ec.europa.eu/projects/lucas
[43] http://54.229.242.119/GSOCmap/
[44] https://www.iuss.org/organisation-people/organisation/working-groups/
[45] GlobalSoilMap.net 2011. Specifications - Version 1 GlobalSoilMap.net products (Release 2.1), 50 p. http://www.globalsoilmap.net/system/files/GlobalSoilMap_net_specifications_v2_0_edited_draft_Sept_2011_RAM_V12.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

2 Current situation of soil data in EJP SOIL: Evaluation of the stocktake results and other surveys on data sources

Authors: Zs. Bakacsi, S. Molnár, K. Farkas-Iványi, M. Fantappiè

2.1 Executive summary EJP SOIL data sources stocktake

This chapter is a summary of the stocktake, in which EJP SOIL partners described the current availability and status of soil properties (SP) and soil management (MG) data. The EJP SOIL partners provided information by completing a questionnaire.

The stocktake concerns not just an overview of the existence of soil data itself, but also auxiliary information needed for soil mapping, and the analytical methods behind the reported databases. During the survey of the measurement methods of the individual soil properties, the stocktake focused on the soil information arranged in extended databases, which does not exclude the possibility that research groups of the individual countries use alternative methods.

From the about 240 reported databases, 210 are georeferenced, 180 databases contain information on soil properties (140) and soil management (90), from which 50 overlap. About 50 other databases were reported as sources for information on climate, vegetation cover, parent material or relief. About 100 soil maps are reported and almost 90 are associated with one of the soil-forming factors (vegetation, parent material, climate). 65 % of the maps have countrywide coverage; the rest of them cover smaller regions or belong to experimental sites.

Long-term experimental farm databases were outside the scope of this stocktake, but some partners reported them as well. Although a single experimental farm dataset is not considered as a countrywide expandable database, the reported sites were not excluded from the final evaluation, because even if it is not spatially representative, the analytical background of the performed measurements (as the other focus of this stocktake) reflects the soil analytical practice of the country and can be useful information for EJP SOIL projects.

The chapter reviews the level of access that countries grant to each database, but does not detail how and under what conditions each country provides free access to data. Free access to soil properties and management data is mainly restricted to derivatives available in different resolutions and often to the content displayed on the web. In general, the access to basic soil properties and especially soil management databases is limited, typically within the remit of national ministries of agriculture.

The most extensive datasets are available for soil properties, such as organic carbon (OC), particle size distribution (sand/silt/clay), gravel content, chemistry (pH water), effective cation exchange capacity (ECEC), bulk density, and calcium carbonate content. The stocktake covers the uptake on a routine basis of new spectral methods. The use of hyperspectral technology using near-infrared (NIR) or mid-infrared (MIR) spectroscopy in soil analysis and survey were reported by Denmark, France, Germany, Italy, the Netherlands, Sweden and Poland. For several questions we received fewer answers compared to the WP2 of EJP SOIL questions (T2.4.2 - soil quality indicators), which may be because the respondents could not assign a measurement methodology to the given indicator and did not reply. The methods used to measure the most common soil properties have been marked on European

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

overview maps, this approach will help to detect the reason of possible transboundary mapping problems.

The soil-management-related national data collection systems are similarly structured, due to central reporting requirements in the EU, while in-country data processing methods and the access levels of each dataset are different. The most frequently available soil management data in reported datasets are: crop type (data for main crop), farming system (conventional/organic/other), fertilization, tillage (conventional/ reduced/ strip-till/ no-tillage), and cropping structure (e.g. monoculture, kind of rotation) data. Access to national soil management databases is limited to different levels, which is regulated by the national ministries of agriculture.

2.2 Introduction

This stocktaking activity aims at collecting metadata information on the georeferenced soil and soil management data available in the EJP SOIL countries. The stocktake focuses on available databases and maps for soil properties and soil management, their spatial and temporal relevance, data collection methods, applied measurements, and data access. This chapter summarizes the general results, details are given in Annexes 2, 3, 4.

2.3 Methodology and source of data

2.3.1 Structure of the questionnaire

This stocktaking activity aims for the provision of a thematic metadata soil information system (SIS) of agricultural soil properties (SP) and management data (MG), for EJP SOIL participating partners. The thematic metadata SIS is considered a baseline, representative of the current condition of the agricultural soil data and encompasses a minimum set of key properties for which there are national and international accounting or reporting obligations (see chapter 1).

The EJP SOIL partners collected the information for this stocktake by completing a questionnaire in excel worksheets, developed by EJP SOIL WP6. This included information on their institutional and often also national soil data holdings. Countries that do not take part in EJP SOIL have not been asked to fill in the questionnaire. The excel-based questionnaire is divided into three main parts (Figure 2.1): 1. general description of data sources, 2. available soil property data (SP), 3. available soil management data (MG). An extensible list of the applied analytical methods, a drop-down list for multiple-choice questions, and some explanatory description of the requested data have been provided in the questionnaire as well. We harmonized the collected management data to EJP SOIL WP2 requirements, as used in their questionnaire (see: T2.4.1. Synthesis of the impacts of sustainable soil management practices[46]).

[46] https://projects.au.dk/knowledge-sharing-platform/reports-publications/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

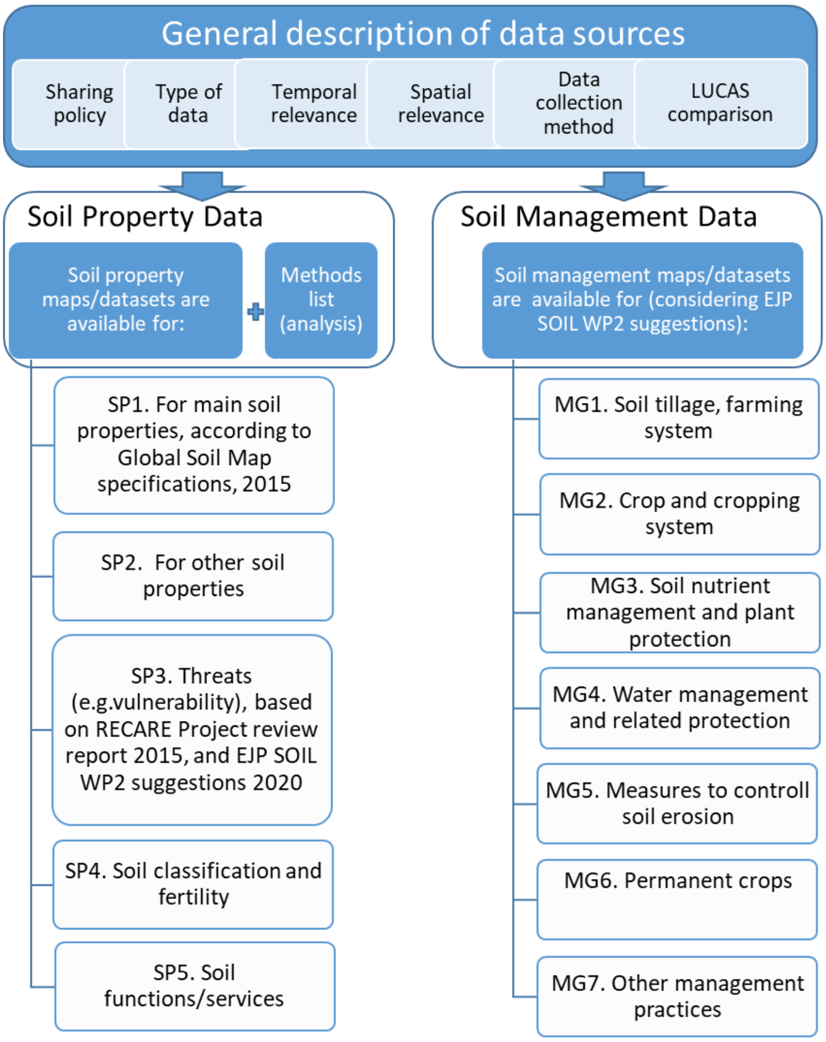


Figure 2.1 Overall structure of the WP6 questionnaire.

2.3.1.1 General description of data sources

25 EJP SOIL partners reported about 240 data sources in the stocktake. 140 databases contain data for soil properties and about 90 contain information for soil management (about half of the management databases contain information for soil properties at the same time) and some other datasets e.g. for climate, forests and geology were also reported. The amount of identified data sources per country are as follows (see the tables in Annex 2, 3, 4):

Austria 7, Belgium Flanders 14, Belgium Wallonia 4, Czech Republic 5, Denmark 7, Estonia 6, Finland 2, France 13, Germany 5, Hungary 5, Ireland 4, Italy 12, Latvia 15, Lithuania 3, Netherlands 12; Norway 15, Poland 6, Portugal 3, Slovakia 5, Slovenia 21, Spain 10, Sweden 15, Switzerland 1, Turkey 14, United Kingdom 35.

The way and level of detail with which the countries report their datasets varies from country to country. There are countries where the database (e.g. point with attributes) and the connecting thematic maps are reported separately (e.g. France, Turkey, Poland). In some cases, the main database and its sub-parts (e.g. profiles, their description, and polygon datasets within one database) are reported separately (Belgium, France, Latvia, Lithuania, the Netherlands, Norway, UK). And there are

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

also cases where an integrated database is reported (in Germany). Despite the fact that the primary objective of this stocktake was to list the metadata of the main national databases in the partner countries, the spatial extent of the reported databases varies from national to local. In some cases partners focused on reporting farm-scale databases (e.g. Spain reports detailed sets of data for experimental sites). A large number of databases within the UK are reported for England, Scotland, Northern Ireland, and Wales separately.

The general description of data sources involved six aspects:

1. Reference and data sharing policy (Database owner, Reference, Sharing policy);

2. Type of data (SCORPAN factor, Proximal/Remote sensed data);

3. Temporal details (beginning-end, frequency of data collection);

4. Spatial details (spatial reference, spatial entity, the geometry of mapping area, the ratio of map cover, spatial resolution);

5. Methods (data collection methodology (sampling, lab methods, mapping methods), depth intervals, uncertainty quantification);

6. LUCAS[47] (processed comparison with LUCAS data).

2.3.1.2 Reference and sharing policy

The reported data policies of the reported datasets are as follows:

About 45% of the reported data sources are freely available (some of them within the EJP SOIL Programme, only for research purposes), almost half of them are available with permission, the rest ("other" category) are partly available (Figure 2.2), for instance for specific purposes. Available means that the data "exists" at a specific data owner (which can be different from the EJP SOIL partner institution) and their use is regulated by a specific sharing policy. The proportion of data that is available "freely" (open data policy) and "with permission" is reversed in the reported soil property (SP) and management (MG) data sources. Data policies for national soil datasets are explored further in deliverable D6.2 (due in July 2021).

The separation of soil data into soil properties and soil management information seems artificial for some data sources because in some countries the database otherwise considered as a soil management dataset which also contains most of the soil properties information (e.g. Germany BZE_LW).

[47] https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

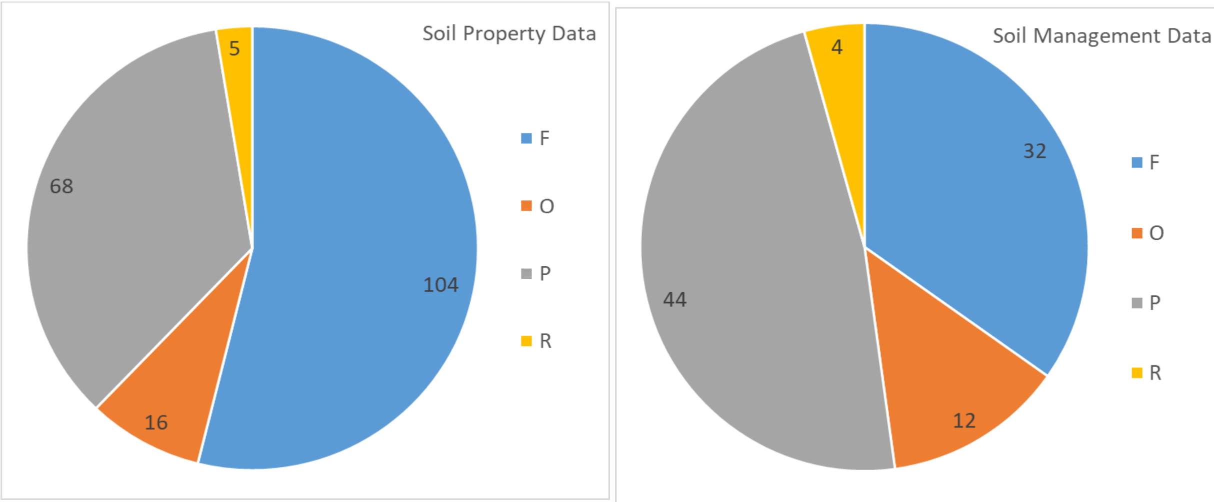


Figure 2.2 Data policy, indicating the reported number of databases by EJP SOIL partners. Data availability as follows: F-freely; R- free for research purposes within EJP SOIL; P-with permission; O- other.

The stocktake does not cover the detailed conditions of open data access, but it seems from the respondents' comments that an open data policy is mainly assigned to derived maps available in different resolutions and content displayed on the web. Access to basic soil property- and soil management data is much more limited, typically within the remit of national ministries of agriculture.

2.3.1.3 Type of data (SCORPAN factor, Proximal/Remote sensing data)

The stocktake also requested to specify the SCORPAN-type of the reported dataset and to assess the proportion of data produced by proximal or remote sensing technology. The SCORPAN is the digital representation of soil forming factors.

McBratney et al. (2003) elaborated the SCORPAN model where soil classes or soil attributes are predicted from the so-called six S,C,O,R,P,A,N factors using a spatial soil prediction function with autocorrelated error: Sc = f(S,C,O,R,P,A,N) + e, or Sa = f(S,C,O,R,P,A,N) + e, where Sc are soil classes; Sa are soil attributes; e is the autocorrelated error; and the six S,C,O,R,P,A,N factors are: **S**: soil, properties of the soil at a point; **C**: climate, climatic properties of the environment at a point; **O**: organisms, vegetation or fauna or human activity; **R**: topography, landscape attributes; **P**: parent material, lithology; **A**: age, the time factor; **N**: space, spatial position.

Almost 160 reported datasets belong to “S” soil factor (70%). The distribution of the remaining data among the other factors is as follows: Climate: 5%, Organisms: 7%, Parent material: 6%, the remaining 12 % are non-specified. The number of reported databases and maps for different SCORPAN factors, based on proximal or remote sensing measurements varies greatly from country to country. Less than half of the partners explicitly report the use of proximal/remote sensed data: Belgium Flanders (2), Belgium Wallonia (4), France (1), Ireland (2), Italy (1), Latvia (1), Norway (13), Poland (1), Sweden (2), United Kingdom (22 for England, Scotland, Northern Ireland, and Wales separately).

2.3.1.4 Temporal relevance (beginning-end, frequency of data collection)

The purpose of these questions was to estimate the temporal validity and detail of the databases. Recent data can describe the actual soil conditions, while earlier (or legacy) data allows deriving or modelling trends or changes over time, or an earlier baseline status.

We examined whether the described databases are the result of single data collection campaign, a regular monitoring survey, or a (in some cases quite long) mapping campaign. The longest monitoring datasets come from *long-term field/farm/site experiments and farm networks*. The primary purpose of the stocktake was not to survey single long-term experimental sites, but some countries also reported these data (Latvia, Spain). The oldest reported long-term field trials "Vecauce" in Latvia has been monitored since 1920. In Spain, the longest reported observation period belongs to "INIA-Experimental Farm La Canaleja (Madrid, Spain)" which has been connected to soil fertility research for 25 years. There are several other long-time experimental farm datasets throughout the EU (e.g. Rothamsted Research – UK - were long-term experiments are running from 1843 until present and in Versailles – F - were experiments on soil fertilization started in 1928), but this issue was outside the scope of the presented report.

The "Specialized project data and long term research database (BAW)"[48] of Austria with more than 1000 observation points over the whole country is 75 years old and is considered to be a farm-based monitoring network. Another extended network of monitoring farms covering a region has been operating in Italy for 12 in private companies.

Apart from the main basic soil properties (like OC, pH, texture) the frequency of specific soil data collection can be unevenly divided in time. In general, the frequency of data collection for monitoring depends on the aim of the monitoring. Within national monitoring systems, according to the monitoring design, the time to return to the same sampling site (point) varies from 1 year to 15 years.

[48] https://www.baw.at/en/wasser-boden-en/projekte.htm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

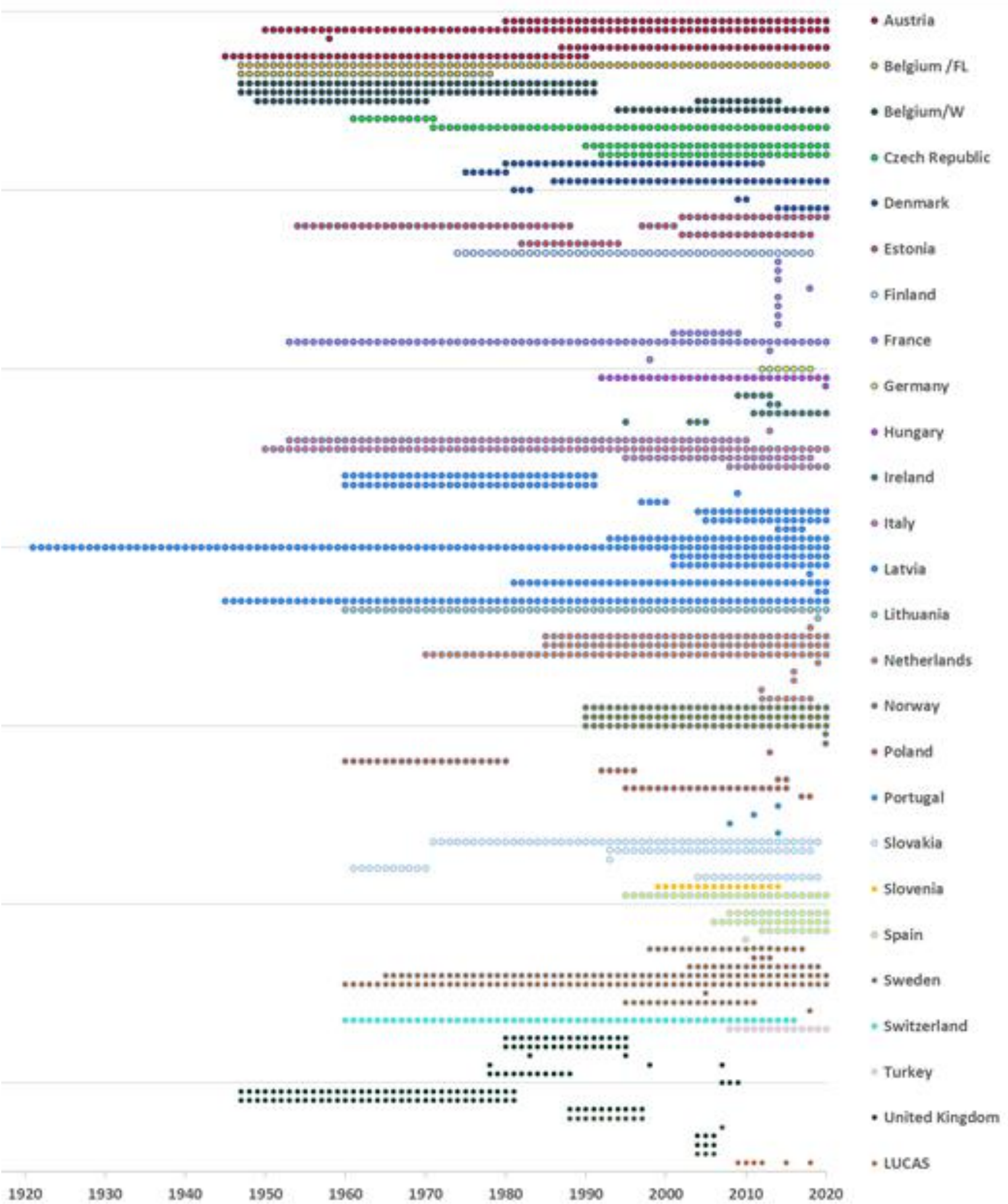


Figure 2.3 Temporal details of the soil property datasets reported by EJP SOIL partners. The unique points represent single surveys, while the chain-linked points refer to monitoring or a mapping campaign.

From Figure 2.3 it is clear that the "traditional" annual national soil sampling is gradually being relegated to the background, because of its especially financial and human resource intensiveness. The data collection density of soil sample-based monitoring networks with less frequent return time (5-10-15y) matches the recurring survey campaigns-like repeating soil inventories (e.g. Germany, France). The spatial pattern of sampling is also changing, while previously the return to "characteristic" points was the main principle, the newly launched monitoring systems with an equidistant network (e.g.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

France and the EU-wide LUCAS monitoring), targeting uniform spatial coverage that better meets the objectives of digital mapping. In some cases, LUCAS monitoring data are already linked to national topsoil monitoring (Latvia, Portugal, and Switzerland).

Figure 2.4 shows the longest period of the ongoing (from the beginning until 2020) national or regional soil property monitoring databases (where the indicated SCORPAN factor is “S”), reported by Partners (Austria, Belgium Wallonia, Czech Republic, Denmark, Estonia, France, Hungary, Latvia, Lithuania, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Switzerland, United Kingdom). As reported by Lithuania their monitoring periodicity is unique, it has no specific time-period fixed. It also happens that a wide range of the soil properties was measured at the beginning of the annual soil monitoring, but the repetition of the same type of measurements can be 3-5 years (Hungary). Due to the different sampling frequency, within the reported regular monitoring systems (shown in Figure 2.4), the youngest available soil properties monitoring data are as follows: Denmark (2019), Estonia (2018), Hungary (2016), Poland (2015), Switzerland (2016), United Kingdom (CEH Topsoil database from 1978, 1998 and 2007). The soil property and management data were included in the first, 8*8 km grid-based German Agricultural Soil Inventory (BZE_LW)[49] which was performed in 2012-2018.

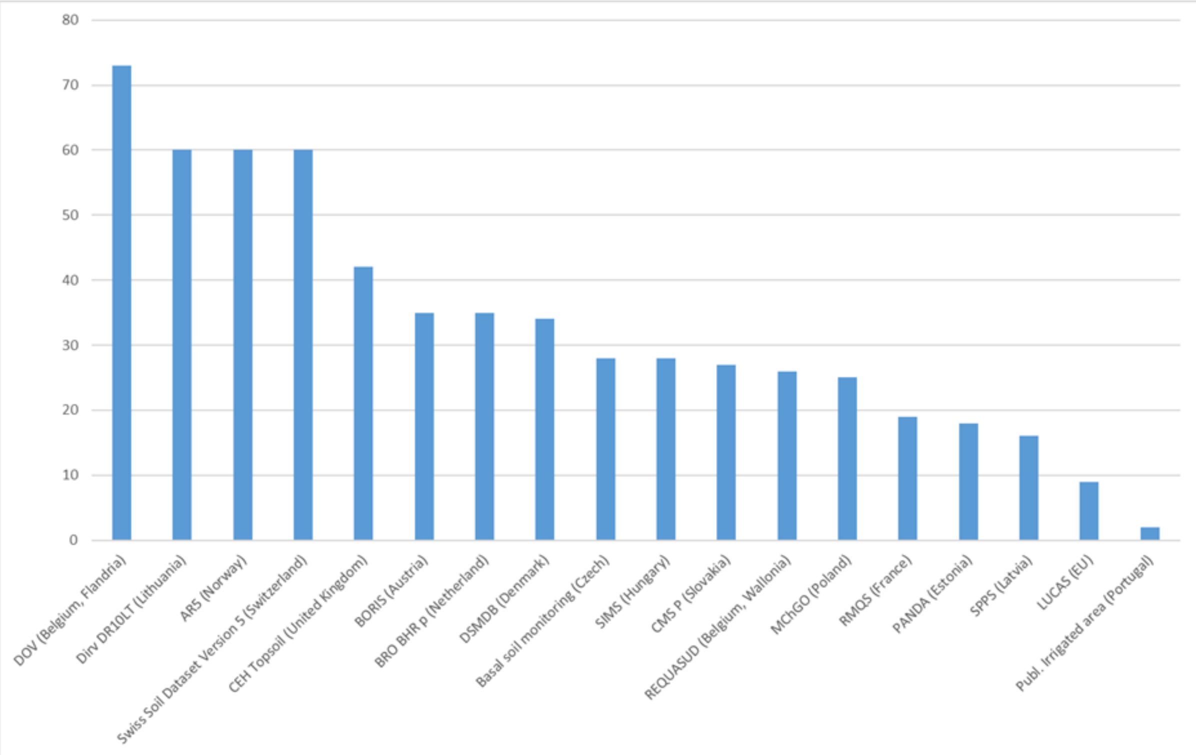


Figure 2.4 Length (Y-axis in years) of the reported, recently working, national and regional soil property monitoring programs, reported as longest regular monitoring by EJP SOIL partners in 2020.

There are fewer reported soil management databases and maps (90) than soil property datasets (140). In 54 cases the soil management and soil property databases partially overlap within the reported data sources (Austria (4), Belgium (3), Czech Republic (4), Denmark (2), Estonia (2), Finland (2), France (2),

[49] https://www.thuenen.de/en/ak/projects/agricultural-soil-inventory-bze-lw/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

Germany (1), Hungary (1), Latvia (9), Lithuania (3), the Netherlands (4), Portugal (1), Slovakia (3), Spain (10), and Sweden (3)).

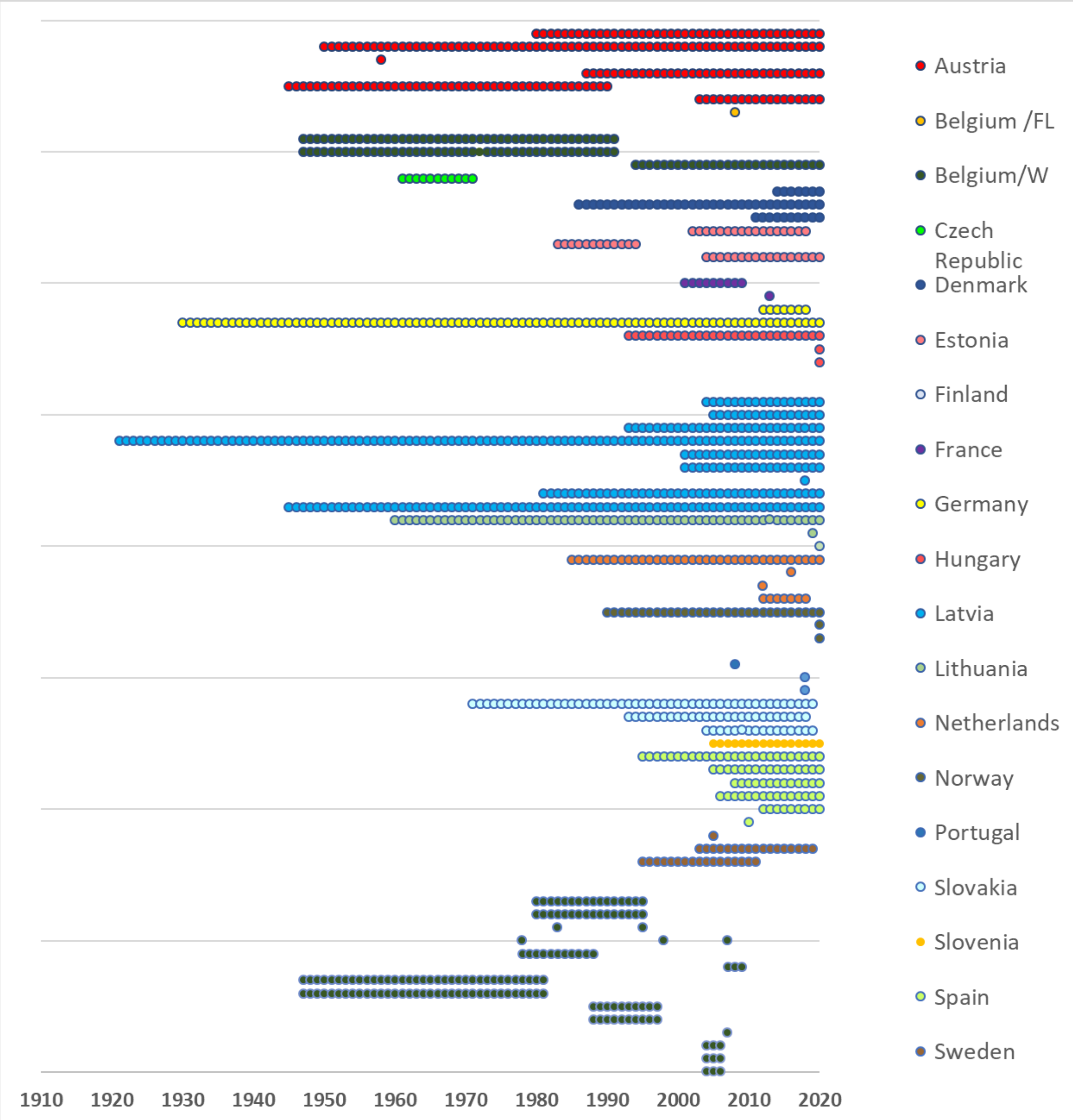


Figure 2.5 Temporal relevance of the soil management datasets reported by EJP SOIL partners. The unique points represent single surveys, while the chain-linked points refer to monitoring or a mapping campaign.

2.3.1.5 Spatial relevance (spatial reference, spatial entity, geometry of the mapping area, ratio of map cover, spatial resolution)

Partners reported predominantly georeferenced data, with almost 210 databases of which 100 are maps. 60 of the reported 100 maps refer to soil properties or soil management data. The non-(or limited) georeferenced datasets mostly contain soil management descriptions; such reported datasets can be e.g. -questionnaires, statistical inventories or farmer's payment registrations.

The most common spatial entities in the reported databases and maps were, in the order of their frequency: points, polygons (with natural boundaries), and grids.

The geometry of the mapped areas partly depends on the purpose of the survey. From the about 100 reported maps almost 80 are associated with one of the SCORPAN factors. 65 % of the maps were spatially exhaustive for the whole country or agricultural lands (their geometry marked as “COUNTRY”). 10 % of the maps cover a spatially exhaustive part of a country (e.g. a region, their geometry marked as “REGIO”), 15 % of the maps cover thematically constrained target areas (e.g. nitrate vulnerable zones, their geometry marked as “TARGET AREAS”) and 10 % remain as non-specified or “other”.

Within the freely available (at least for the EJP SOIL Programme) datasets, the ratio (%) of map cover to the whole agricultural land are slightly different across partner countries (Figure 2.6). Overall, the majority of the maps cover (almost) the entire country. Also a significant part of the data is available as points only.

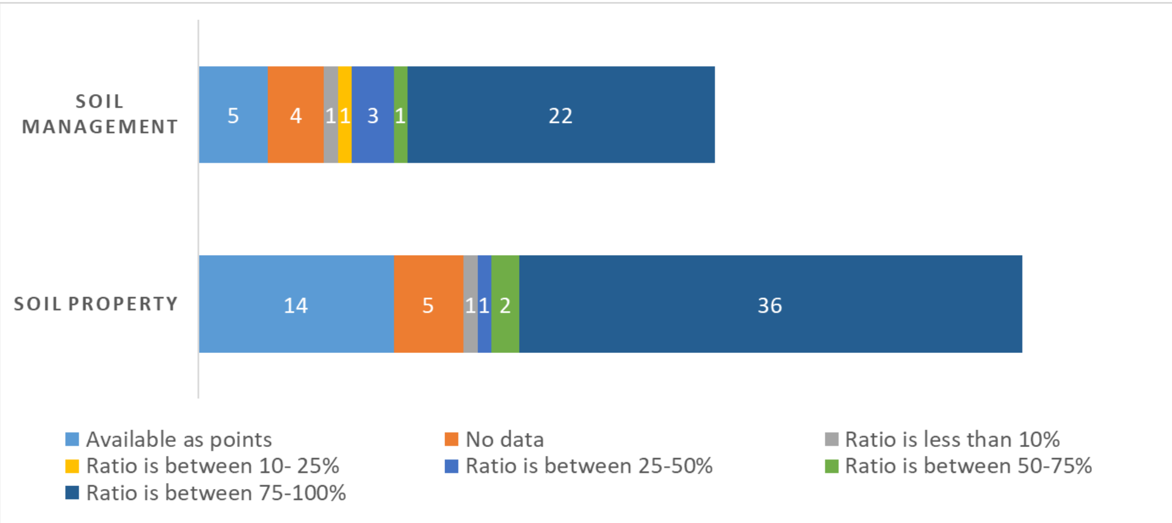


Figure 2.6 Number of datasets (white digit) and the ratio (%) of their coverage compared to the whole agricultural land surface (without datasets only available with special permission), reported by EJP SOIL partners.

About 12 % of all the reported 210 georeferenced databases belong to a network of plots/experiment sites (not considered as maps, their geometry marked as “NETWORK”) or identified farms with management-related unit boundaries (e.g. agricultural parcels). A small number of datasets were reported in which the information was aggregated for administrative polygons (NUT3 level). Figure 2.7 shows the spatial entities and geometry of georeferenced datasets within the whole set of reported data sources.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

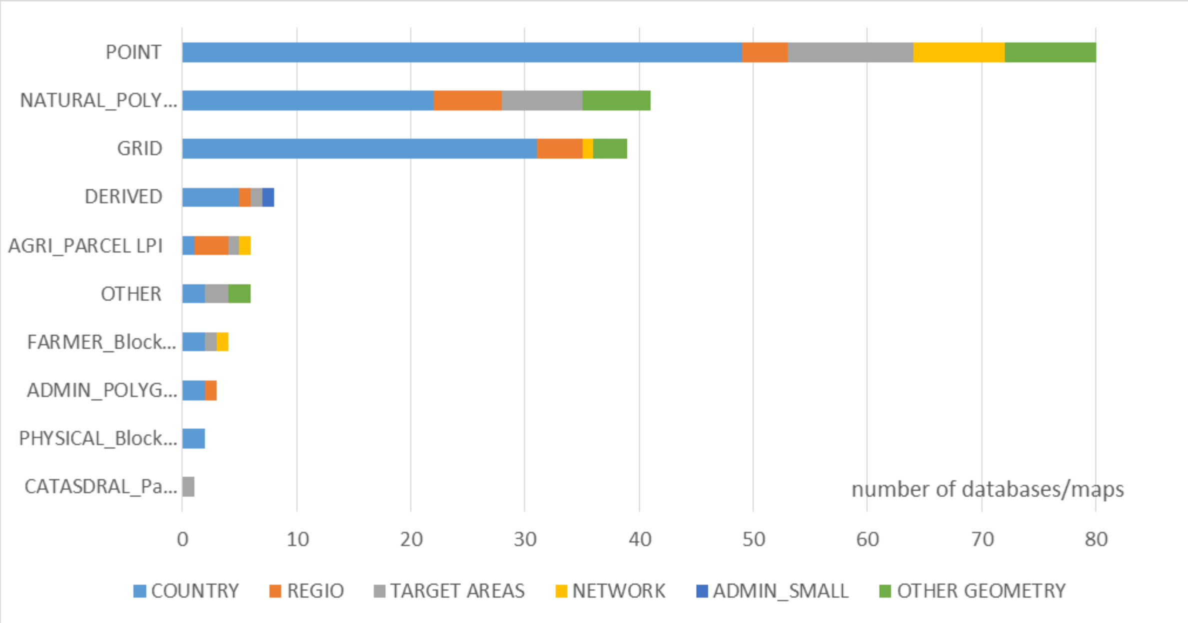


Figure 2.7 Distribution of the spatial entities and geometry of the 210 georeferenced data sources, reported by EJP SOIL partners. POINT: data represented in points; DERIVED (info): soil profiles linked to the polygons; NATURAL_POLYGONs: polygons with natural boundaries; ADMIN_POLYGONS: e.g. municipalities, polygons with administrative boundaries, other than agricultural blocks/farms; AGRI_PARCEL LPI: polygons with LPI are available for agricultural parcel; CATASDRAL_Parcel LPI: polygons with LPI are available for cadastral parcel, FARMER_Block LPI: polygons with LPI are available for farmer's block; PHYSICAL_Block_LPI: polygons with LPI are available for physical block and GRID.

2.3.1.6 Data collection methodology, depth intervals, uncertainty quantification and LUCAS comparison

Data collection methodologies and depth intervals

Traditionally, when soil data is collected in field surveys, the soil can be described in soil pits or boreholes, based on individual or composite soil samples, per horizon or soil depth slice. The sampling depths vary in a wide range (e.g. 0-10 cm, 0-15 cm, 0-20 cm, 0-25 cm, 0-30 cm in "topsoil" databases). Figure 2.8 shows the distribution of soil databases, according to their sampling method as reported by partners. The rates of reported databases with equidistant sampling (with standard depth intervals) and sampling limited for "topsoil" are almost the same. The two-part (topsoil/subsoil) and horizon based sampling are less common strategies. The so-called conventional, soil type based maps often do not contain measured soil parameters or in classified form only, but this kind of database was rarely reported (e.g. United Kingdom).

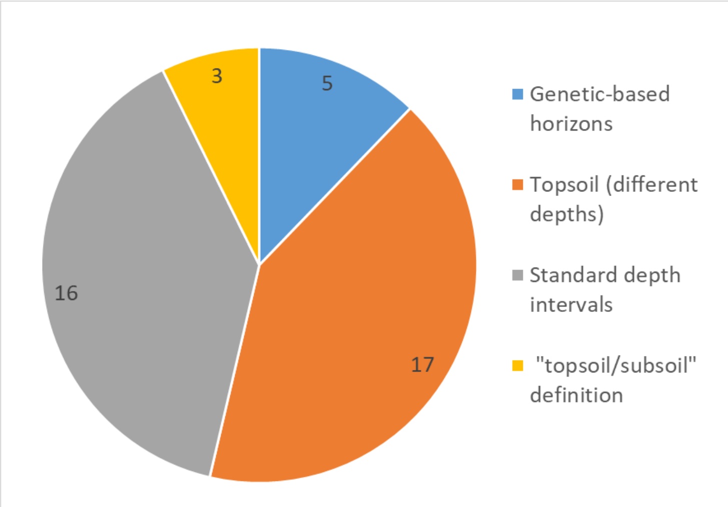


Figure 2.8 : Distribution (and number) of soil databases according to their sampling method, reported by EJP SOIL partners 2020.

Uncertainty quantification

Nine partners have reported that uncertainty analysis was accompanying the databases: Belgium Flanders, Belgium Wallonia, Denmark, France, Germany, Hungary, Ireland, Italy, and the Netherlands.

Comparison of country data with LUCAS

Comparing the information stored in the national Soil Monitoring Networks with the European LUCAS Soil Module data can provide insights in the possibility to use both together for future mapping purposes, for example because together they offer better spatial and soil property coverage of the country. A prerequisite is then to not only compare spatial distribution or mapping patterns, but also possible systematic differences (bias) in the datasets, for instance due to the use of different labs or lab methods, differences in sampling depths, sampling support etc. The main reason for this question in this stocktake is that for further advice on possible harmonization as foreseen in deliverable 6.3 (2nd year of EJP SOIL) it will be valuable to know how partners currently handle the differences (if any) between national and LUCAS datasets. Some partners have already begun, or partially completed the comparison and in some cases, LUCAS has been already integrated into the national monitoring (Latvia, Portugal, and Switzerland). The reported status of comparisons are as follows: Austria (in progress), Hungary (Bakacsi et al., 2020), Estonia (unpublished), France (Mulder et al, 2016), Germany (unpublished), Poland (Łopatka, 2017), Portugal (Ramos et al., 2017), Sweden (unpublished), Italy (in progress). This is further elaborated in EJP SOIL deliverable D6.3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

2.4 Results

2.4.1 Soil Property Data

About 140 soil-property (SP) databases are reported, the amount of identified SP databases per country are as follows as depicted in Table 2.1:

Austria 5, Belgium Flanders 5, Belgium Wallonia 4, Czech Republic 5, Denmark 6, Estonia 4, Finland 2, France 13, Germany 1, Hungary 3, Ireland 4, Italy 5, Latvia 15, Lithuania 3, the Netherlands 12; Norway 5, Poland 6, Portugal 3, Slovakia 5, Slovenia 1, Spain 10, Sweden 9, Switzerland 1, Turkey 1, United Kingdom 14. European datasets are depicted separately (ESDAC).

Table 2.1 Summary of the reported Soil Property Data, where "1" indicates that there is data for that property.

Stocktake of soil thematic maps, point databases and other datasets		Thematic soil **maps/datasets** (SP) are available for:	Austria	Belgium FL	Belgium WALL	Czech Republic	Denmark	Estonia	Finland	France	Germany	Hungary	Ireland	Italy	Latvia	Lithuania	Netherlands	Norway	Poland	Portugal	Slovakia	Slovenia	Spain	Sweden	Switzerland	Turkey	United Kingdom	**sum**	ESDAC (EU)
SP1. for main soil properties, according to Global Soil Map specifications, 2015	SP1.1	total profile depth	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	**23**	1
	SP1.2	plant exploitable (effective) soil depth	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	0	**13**	0
	SP1.3	organic carbon	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**25**	1
	SP1.4	pH in water	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**24**	1
	SP1.5	sand	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**25**	1
	SP1.6	silt	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**25**	1
	SP1.7	clay	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**25**	1
	SP1.8	gravel	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	0	1	**21**	1
	SP1.9	ECEC	1	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**21**	1
	SP1.10	bulk density of the fine earth fraction (excludes gravel)	1	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	0	0	0	0	1	1	1	**12**	0
	SP1.11	bulk density of the whole soil in situ (includes gravel)	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	**19**	1
	SP1.12	available water capacity	1	0	0	0	1	1	0	1	1	1	0	0	0	1	1	0	1	1	0	0	1	0	1	0	1	**13**	1
	SP1.13	Electrical Conductivity	1	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	1	1	0	**15**	1
SP2. for other soil properties	SP2.1	calcium-carbonate content	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	**20**	1
	SP2.2	Field capacity (mm)	1	0	0	0	1	0	0	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	**9**	1
	SP2.3	pH KCl	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	**17**	0
	SP2.4	Saturated hydraulic conductivity (Ksat)	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0	1	1	1	0	0	**8**	1
	SP2.5	Plant available amounts of macro and micro nutrients	1	0	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	0	1	0	1	1	0	0	1	**16**	1
	SP2.6	Total amounts of macro and micro nutrients/trace elem	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	**17**	1
	SP2.7	smectite, montmorillonite in clay fraction…etc)	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**1**	1
	SP2.8	Distribution of soil organisms	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	**6**	1
	SP2.9	precompression stress	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0
	SP2.10	Properties for NIR and MIR (near and mid infrared)	0	0	0	0	1	0	0	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	**7**	1
	SP2.11	*other important, or country-specific soil property maps/da*	1	1	1	1	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	**10**	1
	SP2.12	*other important, or country-specific soil property maps/da*	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	**5**	0
SP3. Threats (e.g.vulnerability), based on RECARE Project review report 2015, and EJP SOIL WP2 suggestions 2020	SP3.1	Soil erosion by water	1	1	0	0	0	1	0	1	0	0	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	**11**	1
	SP3.2	Soil erosion by wind	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	1	**6**	1
	SP3.3	Decline in SOM in peatsoils	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**3**	0
	SP3.4	Decline in SOM in mineral soils	1	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	**10**	1
	SP3.5	Compaction, stucture degradation	1	1	0	0	0	1	0	1	1	0	1	0	1	1	0	0	0	1	1	0	1	1	0	0	0	**12**	1
	SP3.6	Soil sealing	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**2**	0
	SP3.7	Pollution with potentially toxic elements	0	1	0	1	0	1	0	1	0	0	1	0	1	0	0	0	1	1	1	0	1	0	0	1	1	**12**	1
	SP3.8	Pollution with organic substances	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	**7**	0
	SP3.9	Acidification	0	0	0	0	1	1	0	1	0	1	1	0	1	1	0	0	1	1	1	0	0	0	0	1	0	**11**	0
	SP3.10	Salinisation and alkalinisation	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	**4**	1
	SP3.11	Desertification	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1
	SP3.12	Flooding	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	**4**	0
	SP3.13	Landslide	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	**4**	1
	SP3.14	Decline in soil biodiversity	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	**3**	1
SP4. soil classification and fertility	SP4.1	soil type, national classification	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	**23**	0
	SP4.2	soil type, international classification	0	1	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	0	0	1	**17**	1
	SP4.3	soil fertility	0	1	0	1	0	1	1	1	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	**11**	1
	SP4.4	data for status of soil biodiversity/health	0	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	**6**	1
SP5. soil functions/ services	SP5.1	Water storage capacity for topsoil or until a given dept	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	**6**	1
	SP5.2	Plant productivity potential	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	**4**	1
	SP5.3	Carbon sequestration potential	0	0	0	0	0	1	0	1	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	**7**	1
	SP5.4	Biodiversity potential	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	**4**	1
	SP5.5	Other	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	**3**	1
		sum	**24**	**20**	**15**	**19**	**20**	**24**	**15**	**32**	**17**	**27**	**21**	**20**	**28**	**34**	**22**	**16**	**22**	**24**	**25**	**11**	**28**	**22**	**19**	**19**	**24**	**548**	**37**

From the recorded availability of 45 characteristic soil properties (SP) in the datasets/databases, we could subdivide into five groups *(the number of answers from EJP SOIL Project partners is given in brackets)*:

SP1. Main soil properties (according to Global Soil Map specifications, 2015):

- SP1.1 Total profile depth (23)
- SP1.2 Plant exploitable (effective) soil depth (13)
- SP1.3 Organic carbon content (25)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.4 pH in water (24)

SP1.5 Sand content *(indicated the applied particle size intervals) (25)*

SP1.6 Silt content *(indicated the applied particle size intervals) (25)*

SP1.7 Clay content *(indicated the applied particle size intervals) (25)*

SP1.8 Coarse fragments (gravel) (21)

SP1.9 Effective cation exchange capacity (ECEC) (21)

SP1.10 Bulk density of the fine earth (< 2 mm) fraction *(excludes gravel)* (12)

SP1.11 Bulk density of the whole soil in situ *(includes gravel)* (19)

SP1.12 Available water capacity (plant available water) (13)

SP1.13 Electrical conductivity (15)

SP2. Other soil properties:

SP2.1 Calcium-carbonate content (20)

SP2.2 Field capacity (water) (9)

SP2.3 pH KCl (17)

SP2.4 Saturated hydraulic conductivity *(Ksat) (8)*

SP2.5 Plant available amounts of macro and micro nutrients *(16)*

SP2.6 Total amounts of macro and micro nutrients/trace elements *(17)*

SP2.7 Quality of clay minerals *(e.g. type or ratio) (1)*

SP2.8 Distribution of soil organisms *(specified by Partners) (6)*

SP2.9 Precompression stress (0)

SP2.10 Properties for NIR and MIR *(near and mid infrared) (7)*

SP2.11 Other important, or country-specific soil property maps/datasets (10)

SP3. Threats (e.g. vulnerability), based on RECARE Project review report 2015, and EJP SOIL WP2 suggestions 2020:

SP3.1 Soil erosion by water (11)

SP3.2 Soil erosion by wind (7)

SP3.3 Decline in soil organic matter (SOM) in peat soils (3)

SP3.4 Decline in soil organic matter (SOM) in mineral soils (10)

SP3.5 Compaction, structure degradation (12)

SP3.6 Soil sealing (2)

SP3.7 Pollution with potentially toxic elements (12)

SP3.8 Pollution with organic substances (7)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP3.9 Acidification (11)

SP3.10 Salinization and alkalization (4)

SP3.11 Desertification (1)

SP3.12 Flooding (4)

SP3.13 Landslide (4)

SP3.14 Decline in soil biodiversity (3)

SP4. Soil classification and fertility:

SP4.1 Soil type, national classification (23)

SP4.2 Soil type, international classification (17)

SP4.3 Soil fertility (11)

SP4.4 Data for status of soil biodiversity/health (6)

SP5. Soil functions/services:

SP5.1 Water storage capacity for e.g. topsoil or until a given soil depth *(different from SP1.12 or SP2.2) (6)*

SP5.2 Plant productivity potential (4)

SP5.3 Carbon sequestration potential (7)

SP5.4 Biodiversity potential (4)

SP5.5 Other important, or country-specific soil property maps/datasets (3)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

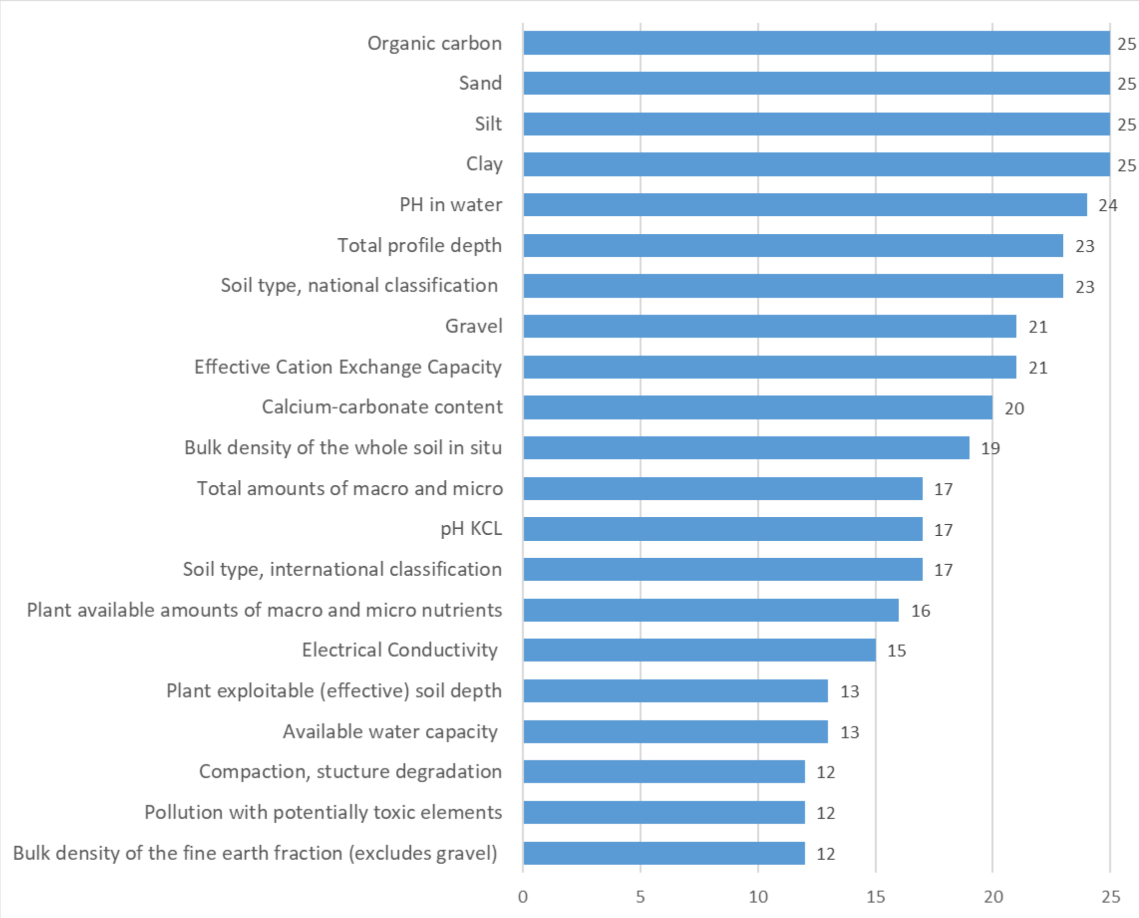


Figure 2.9 Types of frequently reported datasets for soil properties by partners (number of partners positively answered is indicated at bars).

In addition to the available databases for soil properties, the stocktake requires documentation of measurement methods used to obtain the contributors' data. Because of the wide range of collected data and the large variability of measurement methods (including in sample size, sample-management, pre-treatment standards, etc.), some simplification in the description was necessary. The questionnaire included a drop-down menu, which contains about 70 methods often used to determine the given property, with their simplified description. The respondents also had the opportunity to add measurement methods other than those listed. A detailed description of most of the currently used methodologies is available in European HYdropedological Data Inventory (EU-HYDY) with 18682 soil sample data from 18 countries focusing on soil physical, chemical and hydrological properties (Weynants et al., 2013), aimed to develop pedotransfer functions to establish 3D Soil Hydraulic Database of Europe (Tóth et al, 2017).

The assessment of the methods used in the reported soil databases are further complicated by the fact that the measurement and/or sampling methods (depth, design, etc.) may have changed during long monitoring or time-consuming mapping surveys and the database may be heterogeneous in this respect.

The soil properties group SP3 (threats) seems to be underrepresented among reported datasets. The reasons could be that the stocktake is not specified enough in this part, or there is no information about the given property at the EJP SOIL partners, or the derived data has been published as a single

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

map, but not as a part of a larger database. However, it must also be taken into account that not all countries are equally exposed to all threats. Some threats were reported by a few partners only (for soil sealing by Belgium Flanders and Lithuania and for desertification - drought sensitivity, more precisely - by Hungary).

Some soil properties were also exclusively reported by one country (e.g. quality of clay minerals, by Lithuania).

The most frequently available soil properties (at least by 20 Partners) in the reported datasets are:

Organic carbon (OC), Particle size distribution (sand/silt/clay), Chemistry (pH water), Gravel content, Effective cation exchange capacity (ECEC), Bulk density, and Calcium carbonate data. We provide an overview the most frequently reported data below; the other summary tables are in Annex 3.

The stocktake includes the use of spectral methods. The use of near-infrared (NIR) and mid-infrared (MIR) spectroscopy in soil analysis and survey were reported by Denmark, France, Germany, Italy, the Netherlands, Poland and Sweden (Annex 3).

JRC's European Soil Data Centre (ESDAC) provides EU-wide datasets for soil properties, soil threats and functions and more, which may help in further cross-border harmonization excercises (Panagos et al, 2020), foreseen in EJP SOIL WP6 Deliverable 6.3. The datasets are available online[50] with the appropriate metadata after registration.

The list of available ESDAC datasets is as follows:

Soil Point Data

1. LUCAS 2015 TOPSOIL data
2. LUCAS 2009 TOPSOIL data
3. LUCAS 2015 Topsoil data of Switzerland
4. Soil profile analytical database 14 (SPADE 14)
5. Soil Profile Analytical Database 2
6. SPADE/M

Soil Properties and European Soil Database

7. European Soil Database European Soil Database v2.0 (vector and attribute data)
8. European Soil Database v2 Raster Library 1kmx1km
9. SINFO: ESDB Data adapted for the MARS Crop Yield Forecasting System
10. European Soil Database Derived data
11. Topsoil physical properties for Europe (based on LUCAS topsoil data)
12. Maps of Soil Chemical properties at European scale based on LUCAS 2009/2012 topsoil data
13. WRB Data for the Soil Atlas of the Northern Circumpolar Region
14. Soil Atlas of Africa and its associated Soil Map (data)

[50] https://esdac.jrc.ec.europa.eu/resource-type/datasets

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Soil threats datasets

Soil Erosion

15. Soil Loss by Water Erosion in Europe
16. Net erosion and sediment transport using WaTEM/SEDEM (for EU)
17. Soil Erodibility (K- Factor) High Resolution dataset for Europe
18. Rainfall Erosivity in the EU and Switzerland (R-factor)
19. Cover Management factor (C-factor) for the EU
20. Support Practices factor (P-factor) for the EU
21. LS-factor (Slope Length and Steepness factor) for the EU
22. Global Soil Erosion
23. Global Rainfall Erosivity
24. Global soil erosion by water in 2070
25. Global phosphorus losses due to soil erosion
26. Pan European Soil Erosion Risk Assessment - PESERA
27. Soil Erosion Risk Assessment in Europe data (MESALES model)
28. Soil erosion by wind
29. Soil erosion in forestland in Europe (using RUSLE2015)
30. Soil loss due to crop harvesting in the European Union
31. G2 soil erosion model data

Soil Organic Carbon

32. Topsoil Soil Organic Carbon (LUCAS) for EU25
33. Pan-European SOC stock of agricultural soils
34. Carbon budget in the EU agricultural soils
35. Soil Organic Matter (SOM) fractions for 186 LUCAS 2009 soil samples (grassland, forest)
36. OCTOP: Topsoil Organic Carbon Content for Europe
37. Global Soil Organic Carbon Estimates

Soil Biodiversity

38. Potential threats to soil biodiversity in Europe
39. Global Soil Biodiversity Atlas Maps
40. Biodiversity factor in soil erosion

Landslides

41. European Landslide Susceptibility Map version 2 (ELSUSv2)

Salinization

42. Saline and Sodic Soils in European Union

Compaction

43. Natural susceptibility to soil compaction in Europe

Soil Pollution/contamination

44. Copper distribution in topsoils in the European Union
45. Maps of heavy metals in the soils of the EU, based on LUCAS 2009 HM data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

46. Caesium-137 and Plutonium-239+240 in European topsoils

Soil functions

47. Soil Organic Carbon - Saturation Capacity in Europe
48. Maps of indicators of soil hydraulic properties for Europe
49. Maps of the Storing and Filtering Capacity of Soils in Europe
50. Maps of preservation capacity of cultural artefacts and buried materials in soils in the EU
51. European map of soil suitability to provide a platform for most human activities (EU28)
52. Soil Biomass Productivity maps of grasslands and pasture, of croplands and of forest areas in the European Union (EU27)
53. Soil GHG fluxes using LUCAS soil-DayCent (for EU)
54. N2O emissions from agricultural soils in Europe

2.4.1.1 SP1.3 Organic carbon content (OC)

Determination of soil organic carbon (SOC) concentration status is a fundamental part of soil description. It is a widely used basic parameter and common part of indicators targeting soil condition characterization, like soil fertility, soil carbon stocks, and soil health. Because of its importance, it is no coincidence that the Global Soil Partnership chose to map the organic carbon stock first (GSOC map) in a series of indicator maps. About 85% of all Partners (except Hungary and Lithuania) use automated dry combustion method (DRY_ADC), while about 70% of the Partners report that they use one of the wet combustion methods (WB-Walkley Black, Tyurin, or other) to determine OC (Table 2.2 and Figure 2.10).

Table 2.2 Common methods for measuring organic carbon, reported by Partners. The same color of 'yes' indicates the same combination of the applied methods in different countries. Data policy: F-freely (open access); R- free for research purpose EJP SOIL; P-permission; O- other. Applied methods: WET_WALKLEY BLACK: wet combustion: Walkley Black (titrimetric). WET_TYURIN: wet combustion: Tyurin (titrimetric). WET_OTHER: wet combustion: other. DRY_W_LOSS: dry combustion, weight loss on ignition. Sample is heated to 430◦⁰C in a muffle furnace during 24 hours. DRY_ADC: automated dry combustion (CNS, TOC, Elemental Analysis (EA)). Sample is mixed with catalysts or accelerator and heated in resistance or induction furnace in O2 stream to convert all C in CO2.

SP 1.3 Organic C	databases			applied method					
Country	**Relevant for topic**	**data policy**	**(at least a part of it) open access or freely available for EJP SOIL**	**WET_WB**	**WET_TYURIN**	**WET_OTHER**	**DRY_W_LOSS**	**DRY_ADC**	**other**
Austria	4	FPO	eBOD	yes	no	yes	no	yes	no
Belgium Flanders	2	F	DOV, SOCMB	yes	no	no	no	yes	no
Belgium Wallonia	3	P	-	yes	no	no	yes	yes	no
Czech Republic	3	PO	-	yes	no	yes	no	yes	no
Denmark	5	RP	DDSM	no	no	no	yes	yes	no
Estonia	3	FRO	KESE, SMI	no	yes	no	no	yes	no
Finland	2	P	-	no	no	no	no	yes	no
France	4	FP	RMQS, BDAT	no	no	yes	no	yes	no
Germany	1	F	BZE_LW	no	no	no	no	yes	no
Hungary	1	P	-	no	yes	no	no	no	no
Ireland	3	PO	-	yes	no	no	no	yes	no
Italy	5	FP	SISI, PPD, NS	yes	no	yes	no	yes	no
Latvia	7	RP	LLU	no	yes	no	no	yes	no
Lithuania	1	F	DR10LT	no	yes	no	no	no	no
Netherlands	3	O	-	no	no	no	no	yes	yes
Norway	3	FP	NSS	no	no	no	no	yes	yes
Poland	4	FRP	MChGO,MonFrm	no	yes	no	no	yes	no
Portugal	4	FP	INFSOL, PROSOL	yes	no	yes	no	yes	no
Slovakia	2	P	-	yes	yes	no	no	yes	no
Slovenia	1	F	SPSLO	yes	no	no	no	no	no
Spain	5	P	-	yes	no	no	no	yes	no
Sweden	5	FP	SOILCOM	no	no	no	yes	yes	no
Switzerland	1	F	SWISOIL	yes	no	no	no	yes	no
Turkey	1	P	-	yes	no	yes	no	yes	no
United Kingdom	7	FRP	NSI_Top, NSISC88, NSISC09,AFBI 5K, TEL_XRF	yes	no	no	yes	yes	no
%*				52	24	24	16	88	8

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

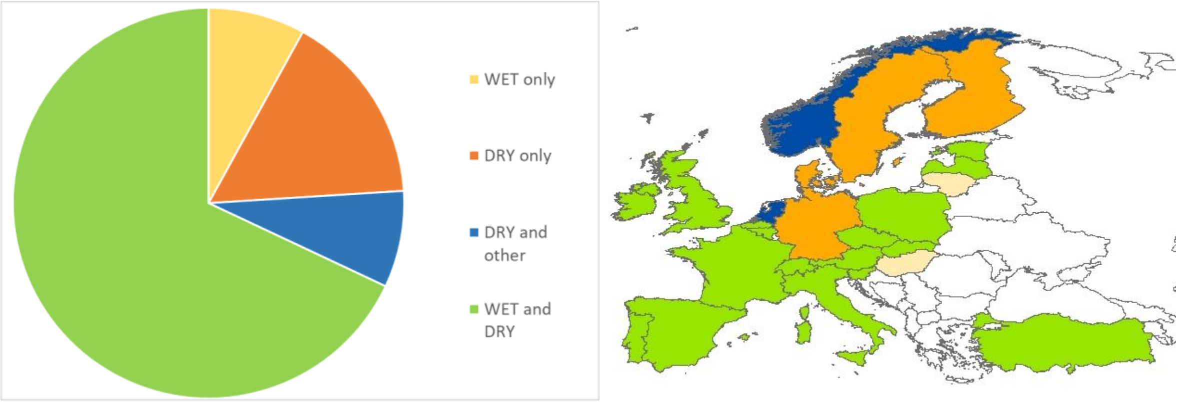


Figure 2.10 Ratio (%, left) and spatial distribution (countries, right) of method's combinations for measuring organic carbon, reported by Partners.

2.4.1.2 SP1.5- SP1.7 Particle size distribution (sand/silt/clay) and SP1.8 Coarse fragment (gravel) content

Particle size distribution (PSD) of solid soil part is a key soil physical parameter. In addition to facilitating classification into texture classes, it refers the soil water-retention characteristics and has a prominent role in pedotransfer functions. About 80% of the Partners applies one of the (sieve)-pipette (sedimentation) method for determining particle size distribution, with or without sample-pretreatment. The various pre-treatment standards (organic matter (OM) and/or $CaCO_3$ and/or iron oxides removal) can modify the PSD results, but they were reported quite rarely. The optical, laser diffraction method (LDM) can be a quick and effective substitute for the time-consuming sedimentation methods. However, its standardization is necessary (Bieganowski et al, 2018) to obtain comparable measurement results; Belgium, Latvia and Poland reported the application of LDM.

Table 2.3 Common methods for determining particle size distribution (sand/silt/clay), reported by Partners. The same color indicates the same combination of the applied pipette methods in different countries. The notation "not specified" indicates that there is a measurement on it, but the method is not reported. Data policy: F-freely (open access); R- free for research purpose EJP SOIL; P-permission; O- other. Applied methods: SIEVE-PIPETTE method, any other details; PIPETTE_NO_PRE: sieve-pipette method no pretreatment; PIPETTE_OM_CACO3_IRON: sieve-pipette method OM, CaCO3 and iron oxides removal; PIPETTE_OM_CACO3: sieve-pipette method OM and CaCO3 removal; PIPETTE_OM: sieve-pipette method OM removal; HYDROMETER METHOD, no details; HYDROM_OM_CACO3: hydrometer method, OM and CaCO3 removal; HYDROM_OM: hydrometer method, OM removal; LDM (laser diffraction method).

SP1.5-7 PSD	databases			size			applied method								
Country	Relevant for topic	data policy	(at least a part of it) open access or freely available for EJP SOIL	sand (mm)	silt (mm)	clay (mm)	sieve-pipette method	PIPETTE_NO_PRE:	PIPETTE_OM_CACO3_IRON	PIPETTE_OM_CACO3	PIPETTE_OM	hydro-meter method	HYDROM_OM_CACO3	LDM	other
Austria	5	FPO	eBOD	0.063-2	0.002-0.063	<0.002	yes	no	no	no	no	no	no	no	yes
Belgium Flanders	1	F	DOV	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	yes	no
Belgium Wallonia	2	P	-	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	no	yes
Czech Republic	4	PO	-	0.05-2	0.001-0.05	<0.001	yes	no	no	no	no	no	no	no	yes
Denmark	3	RP	DDSM	0.02-0.063, 0.063-0.125, 0.125-0.2, 0.2-2	0.002-0.02	<0.002	no	no	yes	no	no	no	no	no	no
Estonia	2	FR	KESE, SMI	0.063-2	0.002-0.063	<0.002	no	no	no	no	yes	no	no	no	no
Finland	2	P	-	0.06-2	0.002-0.06	<0.002	not specified								
France	3	FP	RMQS, BDAT	0.05-2	0.002-0.05	<0.002	no	yes	no	no	no	no	no	no	no
Germany	1	F	BZE_LW	0.063-2	0.002-0.063	<0.002	no	no	no	yes	no	no	no	no	no
Hungary	2	P	-	0.05-2	0.002-0.05	<0.002	no	yes	no	no	no	no	no	no	no
Ireland	2	PO	-	0.05-2	0.002-0.05	<0.002	yes	no	no	yes	no	no	no	no	no
Italy	5	FP	SISI, PPD, NS	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Latvia	7	RPO	DSB	0.063-2	0.002-0.063	<0.002	yes	no	no	yes	no	no	no	yes	no
Lithuania	1	F	DR10LT	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Netherlands	5	FO	BOFEK, BIS-SH	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	yes
Norway	3	FO	NSS	0.06-2	0.002-0.06	<0.002	yes	no	no	no	no	no	no	no	yes
Poland	6	FRP	MChGO, MonFrm	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	yes	yes
Portugal	4	FP	INFSOL, PROSOL	0.02-2	0.002-0.02	<0.002	yes	no	yes	yes	no	no	no	no	no
Slovakia	2	P	-	0.05-2	0.01-0.05	<0.01	yes	no	no	no	no	no	no	no	no
Slovenia	1	F	SPSLO	0.05-0.2, 0.2-2	0.002-0.02, 0.02-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Spain	8	P	-	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	no	no
Sweden	6	FP	SOILCOM	0.06-2	0.002-0.06	<0.002	no	no	no	yes	no	no	no	no	yes
Switzerland	1	F	SWISOIL	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Turkey	1	P	-	0.05-2	0.002-0.05	<0.002	no	no	no	no	no	no	yes	no	no
United Kingdom	4	FRP	NSISC88, NSISC09, AFBI 5K,	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	no	no
%*							68	8	8	20	4	16	4	12	28

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

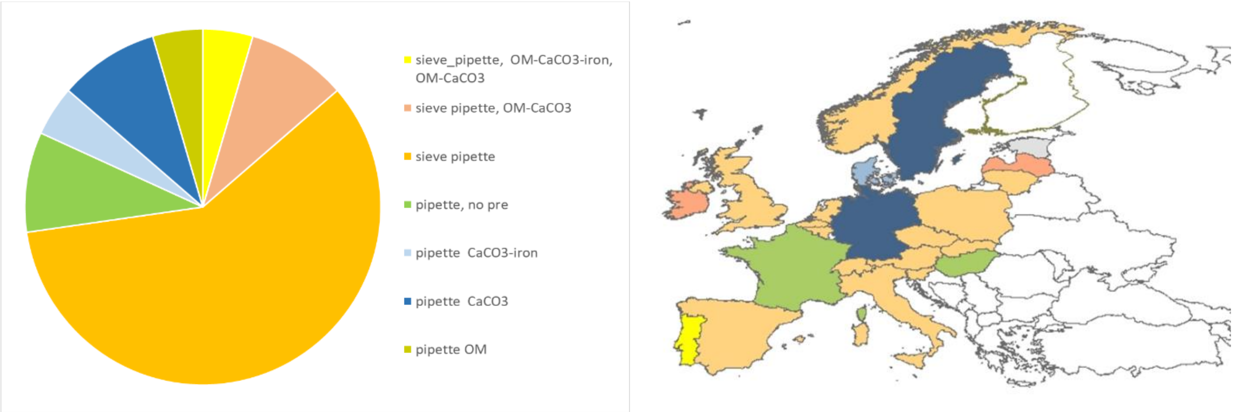


Figure 2.11 Ratio (%, left) and spatial distribution (countries, right) of method's combinations for determining particle size distribution (sand/silt/clay), reported by Partners.

Coarse fragments (gravel) content determination methods are similar in the different countries. Where reported (about 60 of respondents), Partners use the mass-based method, by sieving at 1 or 2 mm (detailed table is in the Annex 3). Coarse fragment determination has an important role in soil water management assessment (e.g. in the mountainous area and/or shallow soils) where the ratio of debris to fine earth material determines the water holding capacity of the soil.

2.4.1.3 SP1.4 pH in water

The acidity or basicity (alkalinity) of a soil (expressed in pH value) is a characteristic soil feature, which is primarily determined by soil-forming processes (like e.g. by bio-chemical soil processes and soil management factors, leaching of carbonates from the profile or accumulation of sodium ions in case of sodic soils), and is therefore one of the sensitive indicators of soil environmental change. About 90% of Partners reported that the pH was primarily measured in soil suspensions in varying solid/water proportions (1:2.5 and 1:5, mostly).

Table 2.4 Common methods for determining pH in water, reported by Partners. The same color indicates the same combination of the applied methods in different countries. The notation "not specified" indicates that there is a measurement on it, but the method is not reported. Data policy: F-freely (open access); R- free for research purpose EJP SOIL; P-permission; O- other. Applied methods: 1:2.5 soil:liquid ratio (water); 1:5 soil:liquid ratio (water); SAT_EXTR: from saturated extract.

SP1.4 pH(water)	databases			applied method			
Country	**Relevant for topic**	**data policy**	**(at least a part of it) open access or freely available for EJP SOIL**	**1:2.5 SLR**	**1:5 SLR**	**SAT_EXTR**	**other**
Austria	4	FPO	eBOD	yes	yes	yes	no
Belgium Flanders	1	F	DOV	yes	yes	no	no
Belgium Wallonia	2	P	-	no	yes	no	no
Czech Republic	2	P	-	yes	yes	no	no
Denmark	3	RP	DDSM	yes	no	yes	no
Estonia	No data	-	-	-	-	-	-
Finland	1	P	-	yes	no	no	no
France	5	FP	RMQS, BDAT	no	yes	no	no
Germany	1	F	BZE_LW	no	yes	no	no
Hungary	2	P	-	yes	no	no	no
Ireland	2	PO	-	yes	no	no	no
Italy	5	FP	SISI, PPD, NS	yes	yes	yes	no
Latvia	2	P	-	no	yes	no	no
Lithuania	1	F	DR10LT	yes	yes	yes	no
Netherlands	3	O	-	no	no	no	yes
Norway	1	P	-	not specified			no
Poland	4	FRP	MChGO,MonFrm	yes	no	no	no
Portugal	4	FP	INFSOL, PROSOL	yes	yes	no	no
Slovakia	2	P	-	yes	yes	no	no
Slovenia	1	F	SPSLO	yes	yes		no
Spain	7	P	-	yes	yes	no	no
Sweden	4	P	-	no	yes	no	no
Switzerland	1	F	SWISOIL	not specified			no
Turkey	1	P	-	yes	yes	yes	no
United Kingdom	7	FRP	NSI_Top, NSISC88, NSISC09,AFBI 5K, UKSHPS, TEL_XRF	yes	yes	no	yes
%*				64	64	20	8

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

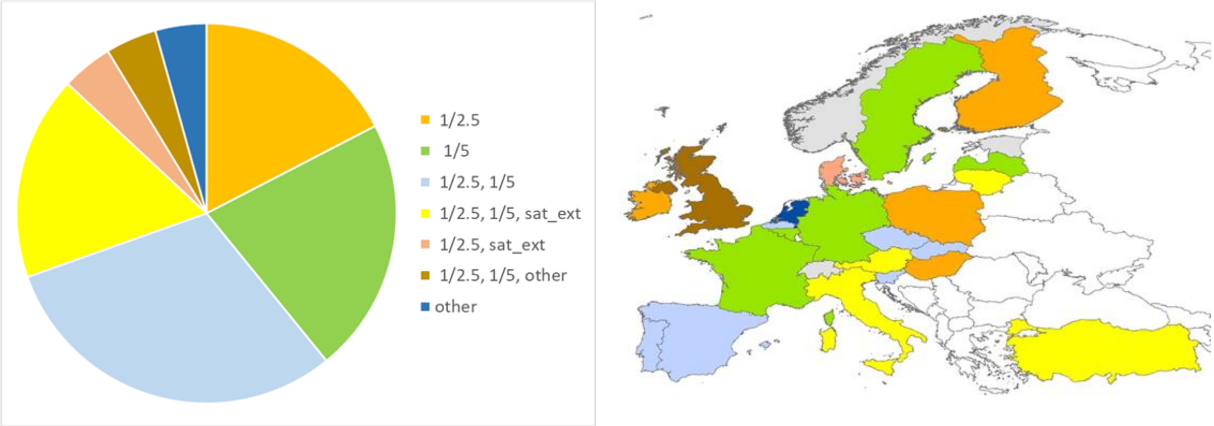


Figure 2.12 Ratio (%, left) and spatial distribution (countries, right) of method's combinations for determining pH in water, reported by Partners.

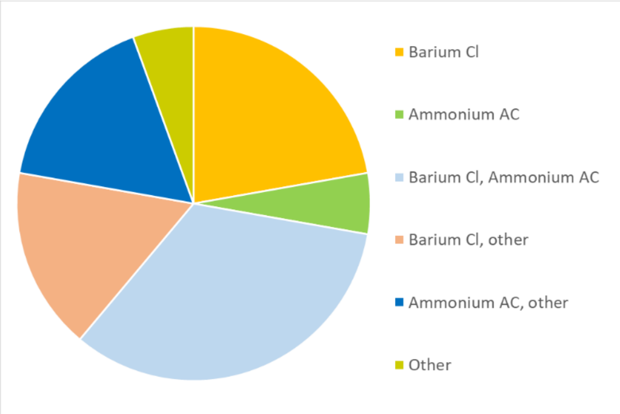
2.4.1.4 SP1.9 Effective Cation Exchange Capacity (ECEC)

About 70 % of Partners reported that the measurement to determine the sum of all exchangeable cations (Na, Mg, K, Ca, and H (and Al)) is mostly obtained by extraction with barium-chloride or by extraction with ammonium acetate. Some other methods were also documented (e.g. spectrometry, flame photometry, or distillation).

Table 2.5 Common methods for determining ECEC, reported by Partners. The same color indicates the same combination of the applied methods in different countries. The notation "not specified" indicates that there is a measurement on it, but the method is not reported. Data policy: F-freely (open access); R- free for research purpose EJP SOIL; P-permission; O- other. Applied methods: BARIUM_CL: Cations extracted using Barium Chloride (BaCl2) plus exchangeable H + Al; AMMONIUM_AC: Cations extracted using Ammonium Acetate plus exchangeable H + Al.

SP1.9 ECEC	databases			applied method			
Country	**Relevant for topic**	**data policy**	**(at least a part of it) open access or freely available for EJP SOIL**	**unit**	**BARIUM_CL**	**AMMONIUM_AC**	**other**
Austria	3	PO	-	mmol IÄ kg-1	yes	yes	no
Belgium Flanders	No data	-	-	-	-	-	-
Belgium Walloni	1	P	-	cmol(+)kg-1	yes	no	no
Czech Republic	3	PO	-	cmolc kg-1, mmol ch. Ekv kg-1, mekv kg-1	yes	no	yes
Denmark	1	P	-	mol 100 g-1	no	yes	yes
Estonia	No data	-	-	-	-	-	-
Finland	No data	-	-	-	-	-	-
France	5	FP	RMQS, BDAT	cmolc kg-1, meq 100g-1	no	yes	yes
Germany	No data	-	-	-	-	-	-
Hungary	1	P	-	cmolc kg-1	yes	yes	no
Ireland	2	PO	-	cmolc kg-1	yes	yes	no
Italy	5	FP	SISI, PPD, NS	cmolc kg-1	yes	yes	no
Latvia	4	P	-	cmol(p+) kg-1, mval 100g-1	yes	no	no
Lithuania	No data	-	-	-	-	-	-
Netherlands	2	O	-	mmol+ kg-1	no	no	yes
Norway	1	P	-	cmolc kg-1	not specified		
Poland	2	FP	MChGO	mmol(+) 100g-1	no	yes	yes
Portugal	4	FP	INFSOL, PROSOL	cmolc kg-1, meq 100g-1	yes	yes	no
Slovakia	2	P	-	cmolc kg-1, mval 100 g-1	yes	no	yes
Slovenia	1	F	SPSLO	mmol 100g-1	yes	no	no
Spain	2	P	-	cmolc kg-1	yes	yes	no
Sweden	1	P	.	cmolc kg-1	yes	no	no
Switzerland	1	F	SWISOIL	mmolckg-1	yes	no	yes
Turkey	1	P	-	cmolc kg-1	yes	yes	no
United Kingdom	1	R	AFBI 5K	meq 100g-1	no	yes	no
%*					56	44	28

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695



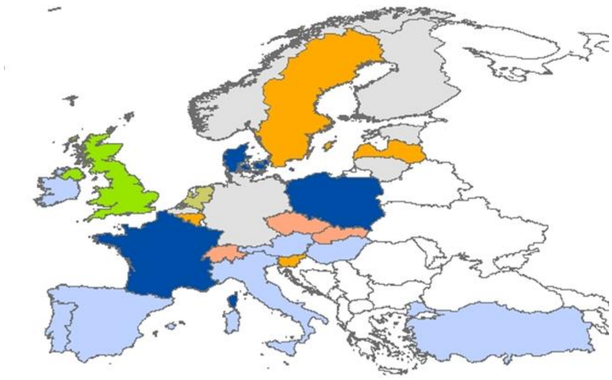


Figure 2.13 Ratio (%, left) and spatial distribution (countries, right) of method's combinations for determining ECEC, reported by Partners.

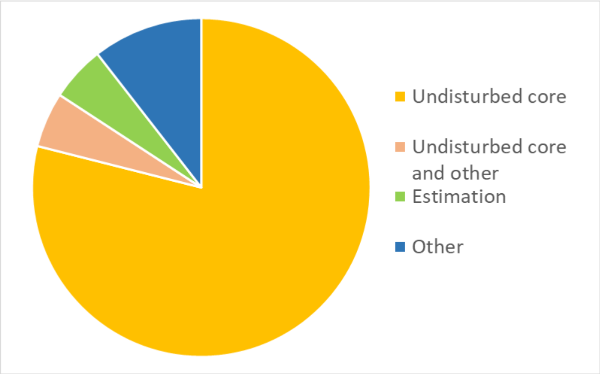
2.4.1.5 SP1.11 Bulk density of the whole soil in situ (includes gravel)

About 64 % of Partners reported that bulk density was predominantly measured by the gravimetric undisturbed core method. However, although the same method was applied some details, like the volume of the applied sample and drying standards (the length of drying at 105°C is usually 48 hours), could be different in countries (Weynants et al., 2013).

Table 2.6 Common methods for determining bulk density, reported by Partners. The notation "not specified" indicates that there is a measurement on it, but the method is not reported. Data policy: F-freely (open access); R- free for research purpose EJP SOIL; P-permission; O- other. Applied methods: UNDIST_CORE: undisturbed core sample; CLODS: coated clods; ESTIMATION: pedotransfer based estimation.

SP1.10-11 Bulk density	databases						bulk density of the fine earth				bulk density of the whole soil		
Country	**Relevant for topic**	**data policy**	**(at least a part of it) open access or freely available for EJP SOIL**	**bulk density of the fine earth**	**bulk density of the whole soil in situ**	**unit**	**ARTIFIC_CORE**	**GRAVEL_CORE**	**EST_**	**calculated**	**UNDIST_CORE:**	**EST_**	**other**
Austria	3	P	-	3	0	dB g cm-3, %	not specified				-	-	-
Belgium Flanders	No data	-	-	0	0	-	-	-	-	-	-	-	-
Belgium Wallonia	No data	-	-	0	0	-	-	-	-	-	-	-	-
Czech Republic	1	P	-	0	1	g cm-3	-	-	-	-	yes	no	no
Denmark	3	RP	DDSM	0	3	g cm-3	-	-	-	-	yes	no	no
Estonia	2	FR	KESE, SMI	0	2	g cm-3	-	-	-	-	no	no	yes
Finland	2	P	-	1	1	g cm-3	not specified				yes	no	no
France	2	FP	RMQS	2	1	g cm-3	no	yes	no	yes	yes	no	yes
Germany	1	F	BZE_LW	1	1	g cm-3	no	no	no	yes	yes	no	no
Hungary	1	P	-	0	1	Mg m-3	-	-	-	-	yes	no	no
Ireland	2	PO	-	0	2	Mg m-3	-	-	-	-	yes	no	no
Italy	5	FP	SISI, PPD, NS	0	5	Mg m-3	-	-	-	-	yes	no	no
Latvia	No data	-	-	0	0	-	-	-	-	-	-	-	-
Lithuania	1	F	DR10LT	0	1	Mg m-3	-	-	-	-	yes	no	no
Netherlands	6	FO	BOFEK	2	4	g cm-3, kg m-3	not specified				yes	no	no
Norway	1	P	-	1	1	-	no	no	yes	no	no	yes	no
Poland	1	R	MonFrm	1	0	-	not specified				-	-	-
Portugal	4	FP	INFSOL, PROSOL	2	2	g cm-3, kg m-3	not specified		no	no	yes	no	no
Slovakia	1	P	-	0	1	g.cm-3	-	-	-	-	yes	no	no
Slovenia	No data	-	-	0	0	-	-	-	-	-	-	-	-
Spain	7	P	-	0	7	Mg m-3, g cm-3	-	-	-	-	yes	no	no
Sweden	1	F	SOILCOM	0	1	g cm-3	-	-	-	-	yes	no	no
Switzerland	1	F	SWISOIL	1	1	g cm-3	not specified				yes	no	no
Turkey	1	P	-	1	0	-	yes	no	no	no	-	-	-
United Kingdom	3	FP	NSISC09, UKSHPS	2	1	t m-3	yes	no	no	no	no	no	yes
%*							8	4	4	8	60	4	12

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695



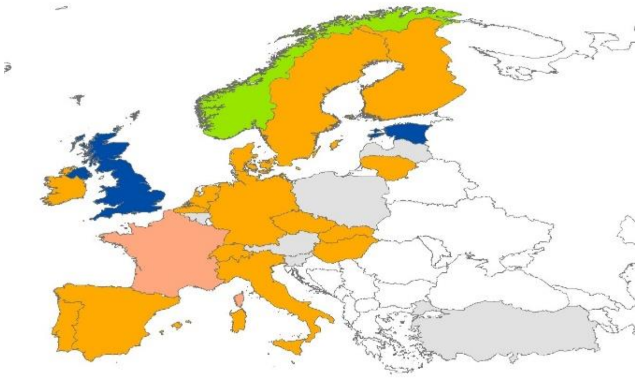


Figure 2.14 Ratio (%, left) and spatial distribution (countries, right) of method's combinations for determining bulk density, reported by Partners.

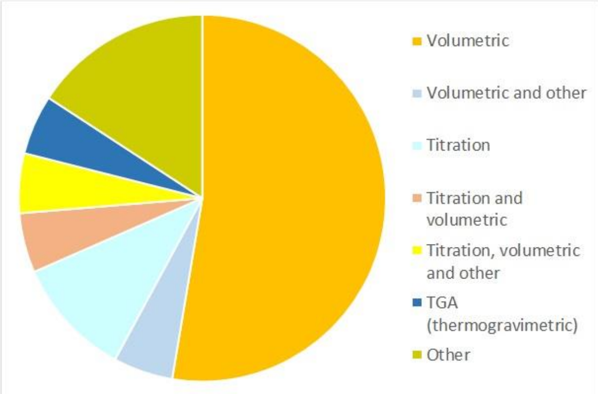
2.4.1.6 SP2.1 Calcium-carbonate content

Most of Partners (52%) reported that they use the volumetric method for calcium carbonate content determination. The remaining values were obtained mostly by titration (16%) or TGA (thermogravimetric) methods.

Table 2.7 Common methods for determining bulk density, reported by Partners. The notation "not specified" indicates that there is a measurement on it, but the method is not reported. Data policy: F-freely (open access); R- free for research purpose EJP SOIL; P-permission; O- other. Applied methods: TITRATION: Titration method (Piper); VOLUMETRIC: with calcimeter (Bernard, Scheibler); ICP: inductively coupled plasma (not reported); XRD: X-ray diffraction (not reported); TGA: thermogravimetric analysis.

SP 2.1 $CaCO_3$	databases				applied method			
Country	**Relevant for topic**	**data policy**	**(at least a part of it) open access or freely available for EJP SOIL**	**unit**	**TITRATION**	**VOLUMETRIC**	**TGA**	**other**
Austria	5	FPO	eBOD	%	no	yes	no	no
Belgium Flanders	1	F	DOV	%	not specified			
Belgium Wallonia	1	P	-	mg kg-1	yes	no	no	no
Czech Republic	2	PO	-	%	yes	yes	no	no
Denmark	3	P	-	%	yes	no	no	no
Estonia	No data	-	-	-	-	-	-	-
Finland	No data	-	-		-	-	-	-
France	3	FP	RMQS, BDAT	g kg-1	no	yes	no	no
Germany	1	F	BZE_LW	g kg-1	no	no	yes	no
Hungary	1	P	-	g kg-1	no	yes	no	no
Ireland	2	PO	-	effervescence	no	no	no	yes
Italy	5	FP	SISI, PPD, NS	g kg-1	no	yes	no	no
Latvia	7	RPO	Dig Prof, LLU	presence, g kg-1, %	no	yes	no	yes
Lithuania	1	F	DR10LT	g kg-1	no	yes	no	no
Netherlands	No data	-	-	-	-	-	-	-
Norway	No data	-	-	-	-	-	-	-
Poland	4	FRP	MChGO,MonFrm	%	no	yes	no	no
Portugal	3	FP	INFSOL, PROSOL	g kg-1, %	no	yes	no	no
Slovakia	2	P	-	g kg-1	no	yes	no	no
Slovenia	No data	-	-	-	-	-	-	-
Spain	1	P	-	g kg-1	no	yes	no	no
Sweden	3	P	-	%	no	no	no	yes
Switzerland	1	F	SWISOIL	%	yes	yes	no	yes
Turkey	1	P	-	g kg-1	no	yes	no	no
United Kingdom	1	P	-	%	no	no	no	yes
%*					16	52	4	20

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695



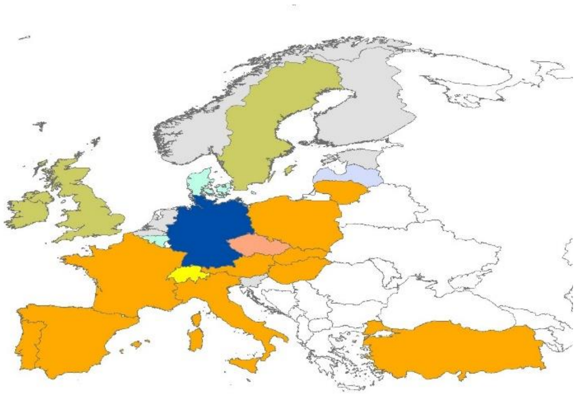


Figure 2.15 Ratio (%, left) and spatial distribution (countries, right) of method's combinations for determining calcium carbonate content, reported by Partners.

2.5 Soil Management Data

2.5.1 Information about agricultural land in EU – short introduction

The provision of agricultural (management) data by EU Member States facilitates the monitoring of agricultural developments through statistics and payments to producers and takes place within integrated frameworks (IACS, Eurostat, Agricultural Census). Therefore the national data collection systems are similarly structured, while in-country data processing is unique (e.g. fully digital or partially analogue farm-data collection, independently managed goal-oriented databases or integrated ones into a common national framework) and the access levels of each data are different.

The WP6 stocktake summarized databases describing management practices, and did not cover the internal data management practices of individual countries other than data policy.

As a common part of the Integrated Administration and Control System (IACS) in the EU, every EU Member State has undertaken the activity to develop a ***Land Parcel Identification System*** (LPIS) to facilitate payments to farmers. The spatially structured LPIS was developed on various data sources, like the Cadastral Register (CR), orthophotos of aerial photographs, topographic maps, and time series of satellite images. Different countries use different levels of land identification (Figure 2.16) such as agricultural and cadastral parcels, farmer and physical blocks (Sagris and Devos, 2008). The national LPISs can also include the development of a GIS of blocks/parcels, integrating area-based information for managing rural development schemes, internet applications together with training for the institutional participants and the clients of the IACS. The renewal and the development of the national LPISs are harmonized with the IACS activities.

				
LPIS types	**Agricultural parcel (AP)**	**Farmer block (FB)**	**Physical block (PB)**	**Cadastral parcel (CP)**
Content/coverage	Single crop group	One or more crop groups	One or more crop groups	Does not match agricultural patterns
Applicant	Single farmer	Single farmer	One or more farmers	One or more farmers
Timeframe	Annual	Multi-annual	Semi-permanent	Not applicable

Figure 2.16 Description of four types of LPIS spatial units within Land Parcel Identification Systems in Europe. According to Sagris and Devos, 2008.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Eurostat[51], the statistical office of the European Union, also collects a wide range of agricultural data on the following main Themes:

1. Farm structure,
2. Economic Accounts for Agriculture,
3. Agricultural prices and price indices,
4. Agricultural production,
5. Organic farming.

Each of the main Themes is divided into several Sub-themes, such as in agricultural products, like: (i) Crop products, (ii) Milk and milk products, (iii) Livestock and meat. On the Eurostat website the aggregated datasets are visualized as tables, graphs and maps. Data related to identification of a statistical unit is treated as confidential and anonymized datasets can be provided for research purposes.

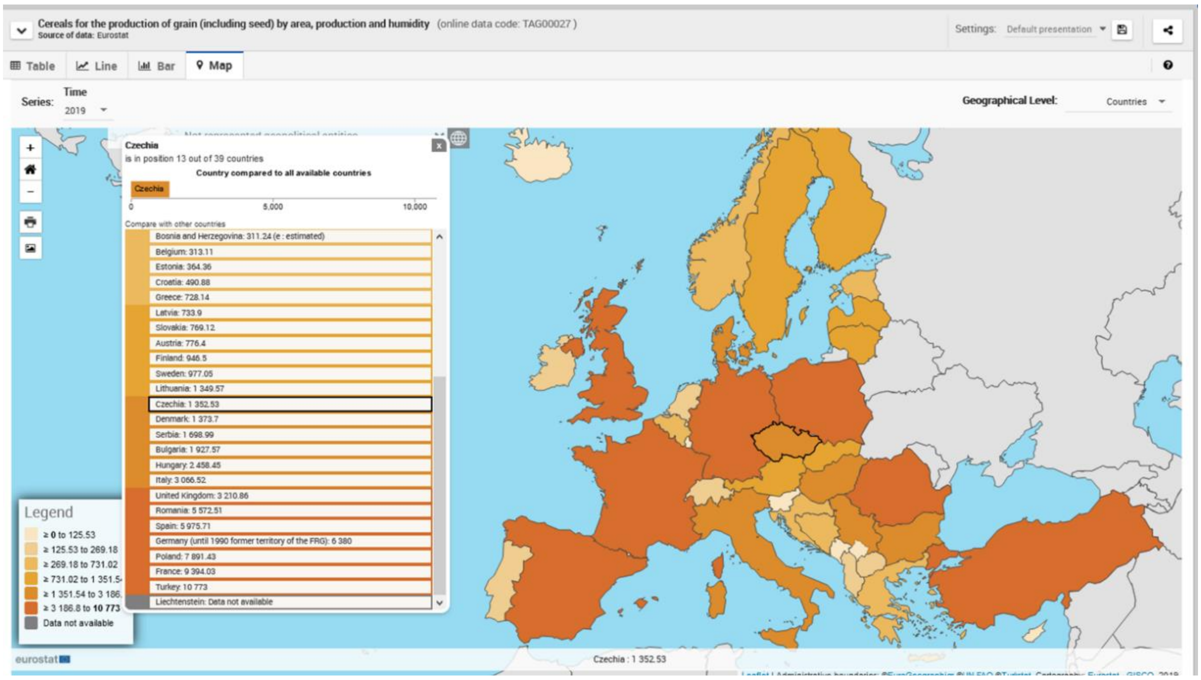


Figure 2.17 Map example from Eurostat web page for "Cereals for the production of grain (including seed) by area (1000 ha) for 2019"aggregated data (https://ec.europa.eu/eurostat/web/main/data/database).

Coordinated by EUROSTAT, the national Statistical Offices conduct comprehensive Agricultural Censuses[52] every 10 years. The decennial Agricultural Census is complemented by sample data collections organized every 3-4 years in-between. The survey aims to monitor the changes in the structure of agriculture and to provide accurate and credible data for national economic governance, the EU, and farmers. In the EU's Agricultural Census 2020 approximately 300 variables were collected, covering the following aspects of farming:

---

[51] https://ec.europa.eu/eurostat

[52] http://www.fao.org/world-census-agriculture/en/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

1. General characteristics of the farm and the farmer
2. Land
3. Livestock
4. Labor force
5. Animal housing and manure management
6. Rural development support measures

The results of the Agricultural Census provide indicators for the monitoring and evaluation of the Common Agricultural Policy[53] (CAP) and the data helps to derive agri-environmental indicators that look at the impact of agriculture on the environment. It seeks answers to questions about how agriculture contributes to positive and negative environmental impacts (soil, air, water, wildlife and climate), how production methods change, or whether agriculture is moving towards a polarized system of both more intensive and more extensive farming.

2.5.2 Stocktake results for Soil Management Data

Land Parcel Identification System (LPIS) is one of the most commonly used georeferenced databases for visualization of collected soil management data. It was possible in the questionnaire to indicate which "spatial entity", the LPIS type (agricultural parcel, cadastral parcel, farmer's block, or physical/topographical block) is available at the Partners, but only eleven Partners indicated it for 17 datasets. (Belgium, Czech Republic, Denmark, Estonia, France, Germany, Hungary, Latvia, Poland, Slovakia, and Sweden).

About 90 soil-management (MG) databases were reported, the number of identified MG databases for a given topic per country are as follows:

Austria 6, Belgium Flanders 4, Belgium Wallonia 3, Czech Republic 2, Denmark 3, Estonia 3, Finland 2, France 2, Germany 2, Hungary 3, Ireland 0, Italy 1, Latvia 9, Lithuania 3, the Netherlands 4; Norway 2, Poland 0, Portugal 3, Slovakia 3, Slovenia 2, Spain 10, Sweden 4, Switzerland 0, Turkey 0, United Kingdom 19.

Data on the same management practice may be included in more than one database (Table 2.8 Summary of the reported Soil Management Data. The numbers for each topic indicate the number of reported databases by countries for given topic.

[53] https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Stocktake of data for soil management systems			Austria	Belgium FL	Belgium WALL	Czech Republic	Denmark	Estonia	Finland	France	Germany	Hungary	Ireland	Italy	Latvia	Lithuania	Netherlands	Norway	Poland	Portugal	Slovakia	Slovenia	Spain	Sweden	Switzerland	Turkey	United Kingdom	SUM
THEMATIC GROUPs		Management practices (MG) – available **data or maps** are about *(concerning WP2 suggestions also)* :																										
MG1. Soil tillage, farming system	MG1.1	Conventional/organic/other farming	3	1	2	1	1	3	2	1	2	2	0	1	6	0	2	1	1	0	1	1	10	4	0	0	1	**46**
	MG1.2	Precision farming	1	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	**5**
	MG1.3	Tillage(conventional/ reduced/ strip-till/ no-tillage)	4	0	1	0	1	2	0	1	2	1	0	0	5	0	3	2	1	0	0	1	10	3	0	0	1	**38**
	MG1.4	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	**2**
MG2. Crop and cropping system	MG2.1	Crop (data for main crop)	4	1	1	1	2	2	2	1	2	3	0	1	5	1	3	1	1	2	2	1	3	4	0	0	3	**46**
	MG2.2	Cropping system (e.g. monoculture, kind of rotation)	4	0	1	0	2	2	0	1	2	0	0	0	6	1	3	0	1	0	1	0	8	3	0	0	0	**35**
	MG2.3	*other, please specify*	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**4**
MG3. Soil nutrient management and plant protection	MG3.1	Fertilization (mineral/organic/both)	3	1	3	0	1	2	0	1	2	2	0	1	7	0	4	1	1	1	1	0	10	2	0	0	2	**45**
	MG3.2	Microbiological preparations	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	**4**
	MG3.3	Pesticides	1	0	2	0	0	1	0	1	0	1	0	1	4	0	0	0	0	0	0	0	2	1	0	0	3	**17**
	MG3.4	*Soil conditioners related to soil /plant health , other than 3.2*	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	**5**
	MG3.5	*other, please specify*	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	1	**5**
MG4. Water management and related protection	MG4.1	Irrigation	3	0	0	0	0	0	0	2	2	1	0	1	2	1	0	0	0	1	0	1	2	0	0	0	1	**17**
	MG4.2	Drainage	2	0	1	1	2	3	0	2	2	0	0	0	4	1	0	0	0	0	0	0	0	2	0	0	0	**20**
	MG4.3	Soil conditioners related to water protection	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	MG4.4	*other, please specify*	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	7	**9**
MG5. Measures to control soil erosion	MG5.1	Terraces (wall, bench, ridges or raised beds, others)	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**3**
	MG5.2	Windbreak hedges	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	**3**
	MG5.3	Runoff water management systems (channels, etc)	0	0	2	0	0	0	0	1	1	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	**7**
	MG5.4	Buffer strips	0	1	2	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	3	0	0	0	**9**
	MG5.5	In field erosion control measures (e.g. micro-dams between	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**
	MG5.6	Small scale buffering infrastructure (retention ponds, dams)	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**2**
	MG5.7	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**2**
MG6. Permanent crops SSM	MG6.1	Green manuring	0	0	2	0	0	0	0	1	2	1	0	0	2	0	2	0	1	0	0	0	1	2	0	0	0	**14**
	MG6.2	Cover crops	1	1	0	0	0	0	2	1	2	1	0	0	2	0	1	0	1	0	1	1	0	3	0	0	0	**17**
	MG6.3	Mulching	0	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	1	0	0	0	1	0	0	0	0	**6**
	MG6.4	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**
MG7. Other management practices	MG7.1	Amelioration (other than drainage M	0	0	0	0	0	0	0	0	1	0	0	1	0	2	0	0	0	0	0	0	0	2	0	0	0	**6**
	MG7.2	Greenhouses	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	**5**
	MG7.3	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**
		SUM	**28**	**9**	**20**	**3**	**9**	**16**	**6**	**17**	**23**	**14**	**0**	**7**	**56**	**12**	**18**	**6**	**8**	**4**	**9**	**5**	**54**	**31**	**0**	**0**	**22**	**377**

). For detailed results of the soil management stocktake, see the Tables in Annex 4.

Table 2.8 Summary of the reported Soil Management Data. The numbers for each topic indicate the number of reported databases by countries for given topic.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Stocktake of data for soil management systems																												
THEMATIC GROUPs	Management practices (MG) – available **data or maps** are about *(concerning WP2 suggestions also)* :		Austria	Belgium FL	Belgium WALL	Czech Republic	Denmark	Estonia	Finland	France	Germany	Hungary	Ireland	Italy	Latvia	Lithuania	Netherlands	Norway	Poland	Portugal	Slovakia	Slovenia	Spain	Sweden	Switzerland	Turkey	United Kingdom	**SUM**
MG1. Soil tillage, farming system	MG1.1	Conventional/organic/other farming	3	1	2	1	1	3	2	1	2	2	0	1	6	0	2	1	1	0	1	1	10	4	0	0	1	**46**
	MG1.2	Precision farming	1	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	**5**
	MG1.3	Tillage(conventional/ reduced/ strip-till/ no-tillage)	4	0	1	0	1	2	0	1	2	1	0	0	5	0	3	2	1	0	0	1	10	3	0	0	1	**38**
	MG1.4	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	**2**
MG2. Crop and cropping system	MG2.1	Crop (data for main crop)	4	1	1	1	2	2	2	1	2	3	0	1	5	1	3	1	1	2	2	1	3	4	0	0	3	**46**
	MG2.2	Cropping system (e.g. monoculture, kind of rotation)	4	0	1	0	2	2	0	1	2	0	0	0	6	1	3	0	1	0	1	0	8	3	0	0	0	**35**
	MG2.3	*other, please specify*	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**4**
MG3. Soil nutrient management and plant protection	MG3.1	Fertilization (mineral/organic/both)	3	1	3	0	1	2	0	1	2	2	0	1	7	0	4	1	1	1	1	0	10	2	0	0	2	**45**
	MG3.2	Microbiological preparations	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	**4**
	MG3.3	Pesticides	1	0	2	0	0	1	0	1	0	1	0	1	4	0	0	0	0	0	0	0	2	1	0	0	3	**17**
	MG3.4	*Soil conditioners related to soil /plant health , other than 3.2*	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	**5**
	MG3.5	*other, please specify*	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	1	**5**
MG4. Water management and related protection	MG4.1	Irrigation	3	0	0	0	0	0	0	2	2	1	0	1	2	1	0	0	0	1	0	1	2	0	0	0	1	**17**
	MG4.2	Drainage	2	0	1	1	2	3	0	2	2	0	0	0	4	1	0	0	0	0	0	0	0	2	0	0	0	**20**
	MG4.3	Soil conditioners related to water protection	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	MG4.4	*other, please specify*	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	7	**9**
MG5. Measures to control soil erosion	MG5.1	Terraces (wall, bench, ridges or raised beds, others)	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**3**
	MG5.2	Windbreak hedges	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	**3**
	MG5.3	Runoff water management systems (channels, etc)	0	0	2	0	0	0	0	1	1	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	**7**
	MG5.4	Buffer strips	0	1	2	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	3	0	0	0	**9**
	MG5.5	In field erosion control measures (e.g. micro-dams between	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**
	MG5.6	Small scale buffering infrastructure (retention ponds, dams)	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**2**
	MG5.7	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**2**
MG6. Permanent crops SSM	MG6.1	Green manuring	0	0	2	0	0	0	0	1	2	1	0	0	2	0	2	0	1	0	0	0	1	2	0	0	0	**14**
	MG6.2	Cover crops	1	1	0	0	0	0	2	1	2	1	0	0	2	0	1	0	1	0	1	1	0	3	0	0	0	**17**
	MG6.3	Mulching	0	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	1	0	0	0	1	0	0	0	0	**6**
	MG6.4	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**
MG7. Other management practices	MG7.1	Amelioration (other than drainage M	0	0	0	0	0	0	0	0	1	0	0	1	0	2	0	0	0	0	0	0	0	2	0	0	0	**6**
	MG7.2	Greenhouses	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	**5**
	MG7.3	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**
		SUM	**28**	**9**	**20**	**3**	**9**	**16**	**6**	**17**	**23**	**14**	**0**	**7**	**56**	**12**	**18**	**6**	**8**	**4**	**9**	**5**	**54**	**31**	**0**	**0**	**22**	**377**

Regarding the existence of soil management (MG) datasets, we discerned five management categories (also considering the EJP SOIL WP2 suggestions for management data), the number of EJP SOIL Project partners that answered positively is given in brackets:

MG1. Soil tillage, farming system:

MG1.1 Conventional/organic/other farming (20)

MG1.2 Precision farming (4)

MG1.3 Tillage (conventional/ reduced/ strip-till/ no-tillage) (15)

MG1.4 Other *(specified by Partners) (2)*

MG2. Crop and cropping system:

MG2.1 Crop (data for main crop) (22)

MG2.2 Cropping system (e.g. monoculture, kind of rotation etc.) (13)

MG2.3 Other *(specified by Partners) (3)*

MG3. Soil nutrient management and plant protection:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG3.1 Fertilization (mineral/organic/both) (18)

MG3.2 Microbiological preparations (3)

MG3.3 Pesticides (10)

MG3.4 Soil conditioners related to soil /plant health *(other than MG3.2) (2)*

MG3.5 Other *(specified by Partners) (4)*

MG4. Water management and related protection:

MG4.1 Irrigation (11)

MG4.2 Drainage (10)

MG4.3 Soil conditioners related to water protection (0)

MG4.4 Other *(specified by Partners) (3)*

MG5. Measures to control soil erosion:

MG5.1 Terraces (wall, bench, ridges or raised beds, others) (3)

MG5.2 Windbreak hedges (3)

MG5.3 Runoff water management systems (channels, etc.) (5)

MG5.4 Buffer strips (6)

MG5.5 In field erosion control measures (e.g. micro-dams between ridges, etc.) (2)

MG5.6 Small scale buffering infrastructure (retention ponds, dams, etc.) (2)

MG5.7 other, please specify (1)

MG6. Permanent crops:

MG6.1 Green manuring (8)

MG6.2 Cover crops (12)

MG6.3 Mulching (4)

MG6.4 Other *(specified by Partners) (0)*

MG7. Other management practices:

MG7.1 Amelioration (other than drainage MG4.2) (5)

MG7.2 Greenhouses (5)

MG7.3 Other *(specified by Partners) (1)*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

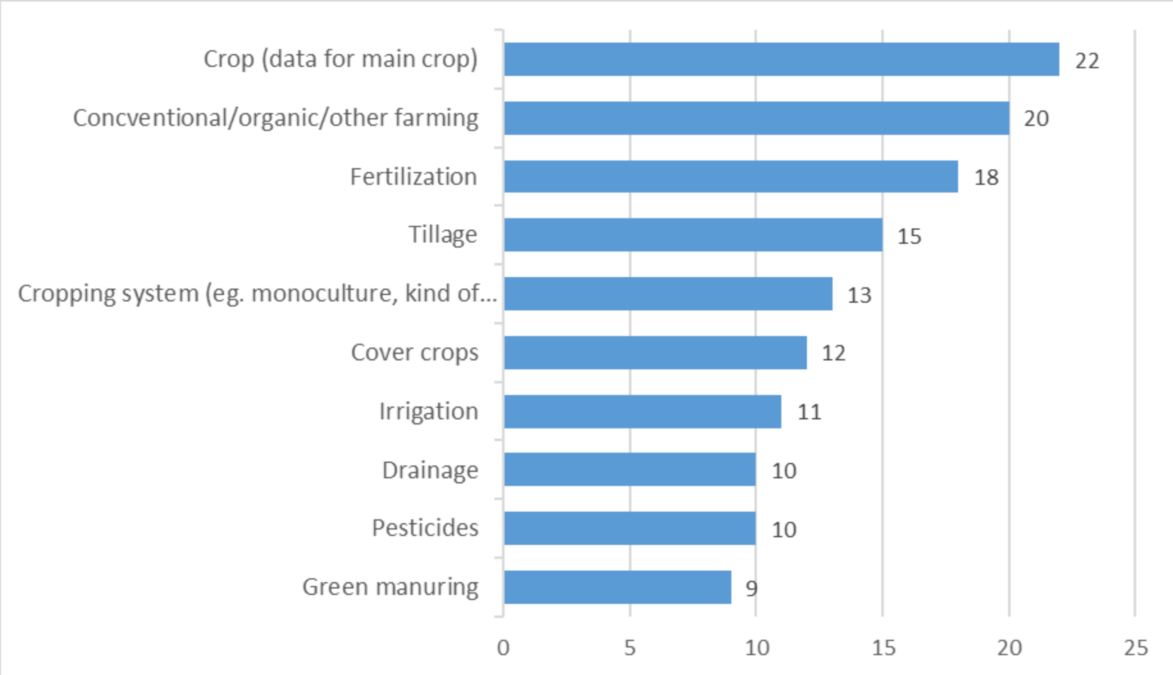


Figure 2.18 Types of frequently reported datasets for soil management (MG) by Partners (number of Partners positively answered is indicated at bars).

In addition to the most commonly reported soil management datasets, about half of the responding partners reported having a database for cover crops, water management or related protection (like irrigation, drainage) and pesticides.

The most frequently available soil management data in the reported datasets are:

1. Crop (data for main crop), 2. Conventional/organic/other farming, 3. Fertilization, 4. Tillage (conventional/reduced/strip-till/no-tillage), and 5. Cropping system (e.g. monoculture, kind of rotation etc.) data. Thanks to the common data framework, incorporation of collected soil management data in the LPISs is similar in different countries. All or part of the management data provided by farmers will be registered in the LPISs.

2.5.2.1 MG2.1 Crop (data for main crop)

About 80% of Partners reported available databases for main crop, with 45 total number of datasets (see detailed Table in Annex 4). Annual data on the crop that is grown are provided by farmers in the form of various questionnaire-based data submitted on digital or analogue basis. For quality assurance, the input provided by farmers is checked via remote sensing.

2.5.2.2 MG1.1 Conventional/organic/other farming

About 70% of Partners reported available databases for farming systems, with 44 total number of datasets (see detailed Table in Annex 4).

2.5.2.3 MG3.1 Fertilization (mineral/organic/both)

About 70% of Partners reported available databases for application of fertilizers with 44 total number of datasets (see detailed Table in Annex 4). Most of the reported databases refer to data provided by farmers in nitrate sensitive areas.

2.5.2.4 MG1.3 Tillage (conventional/ reduced/ strip-till/ no-tillage)

About 50% of Partners reported available databases for soil tillage, with 37 total number of datasets (see detailed Table in Annex 4).

2.5.2.5 MG2.2 Cropping system (e.g. monoculture, kind of rotation etc.)

About 50% of Partners reported available databases for cropping structure, with 34 total number of datasets (see detailed Table in Annex 4).

2.6 Report on CountrySIS survey

The Global Soil Partnership (GSP, see paragraph 1.2.9) has launched a questionnaire to collect information on soil data available at country level, called the CountrySIS survey[54] in 2018. The survey is described by GSP as:

"The survey aims to assess soil databases and information systems currently existing on the national level, in order to plan global activities according to the capacities and needs of the countries. The PDF of the survey questions can be accessed online[55]. The survey starts with questions related to the soil property database. If the database is part of a soil information system (SIS), it then asked to describe the SIS (if not, the SIS section can be skipped). The last section is dedicated to the presence and description of the national soil monitoring system. A soil property database is a collection of measured values of soil properties organized in a digital format so that it can be easily accessed, managed and updated. The data may be associated with soil profiles (soil profile database) or with the mapping units. A SIS is a geographic information system (GIS to capture, storage, management, process and display of soil-related data from original sources. A soil property database may serve as the main component of the SIS. A SIS can also include non-soil data (such as climate or land use) to support land-management decision making. A Soil Monitoring System is based on regular repetitive soil sampling aimed at observing the change of soil properties over time in order to control soil quality and address soil degradation."

In the first months of EJP SOIL WP6 activity and after asking permission to GSP, we have distributed the CountrySIS questionnaire to all EJP SOIL WP6 national representatives. Some of them had already answered to the GSP questionnaire in 2018.

17 countries (Austria, Belgium, Denmark, France, Finland, Germany, Hungary, Italy, Latvia, Lithuania, the Netherlands, Norway, Poland, Slovakia, Spain, Switzerland, Turkey, UK) of the EJP SOIL consortium have answered in 2020 and 2021, and the 2020 answers were given by the EJP SOIL partner institutions (for Switzerland it was given by the Swiss Federal Office for the Environment, after the invitation by AGS, the Swiss EJP SOIL partner). Among these, Spain answered but just to say that they do not have a soil database, neither a SIS nor a monitoring system at national level. For Estonia, Portugal, and Slovenia we have had answers since 2018, but of institutions not belonging to EJP SOIL. For Belgium, Germany, Hungary, Italy, Spain, and United Kingdom we have both the answer from the EJP SOIL institutions and from other institutions. In this report we are publishing only the answers received from EJP -SOIL institutions.

For detailed results of CountrySIS, see the Annex 5. Here we summarize some main results. Summary results of the answers related to the **soil property database** are given in the Table 2.9. The following countries have a **soil information system (SIS)** which is functioning at present (verified on 03/03/2021):

---

[54] https://forms.gle/X6N2G4WX86VYk8tn9

[55] https://drive.google.com/open?id=1rboHEXP-7LO3mY9fFhivk3Gc80lcqBGa

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Austria[56], Belgium[57], Germany[58], Hungary[59], Lithuania[60], the Netherlands[61], Slovakia[62], and Switzerland[63].

Italy declared in 2018 to have a national SIS[64] but it is not active at the verification date. Although the national soil portal of Italy is no more active, several sub-national soil portals are available and maintained by the subnational soil services[65].

The following countries declared that they do not have a SIS: Denmark, Finland, Latvia, Norway, Turkey, United Kingdom (Northern Ireland). A study commissioned by DG-ENV in 2019 "Collection of Meta-Data on Digital Soil Data in Europe; development of a database on Digital Soil Resources of Europe" or Mensmeu 1.0 (data available from ESDAC at ec-esdac@ec.europa.eu), and provided to EJP SOIL by JRC and with permission of DG-ENV, indicates that Czech Republic[66] and France[67] have an active geoportal (verified on 07/03/2021). Other interesting information derived from Mensmeu 1.0, is the availability for several countries of WMS/WCS/WFS services. The countries with WMS/WCS/WFS services (verified active on 07/03/3021) are: Belgium, Czech Republic, Germany, France, Latvia, the Netherlands, Poland, and the UK. No active WMS/WCS/WFS services were given for: Estonia, Finland, Hungary, Lithuania, Sweden, Slovenia, and Slovakia.

Mensmeu 1.0 (Metadata of national soil maps of the European Union) is a result of a study conducted by JRC and DG-ENV to create a database with INSPIRE compliant meta-data for national soil services of Europe focusing on the data available on the web. As the thematic range of such data and their location or the contact details can also change, it is necessary to revise them from time to time. EJP SOIL Project WP6 has asked the partners to update the part of the Mensmeu that applies to them. Some EJP SOIL Project partners have renewed or completed the previously reported data (Estonia, French, Germany, Hungary, Slovakia). In addition to minor changes, the revised database was expanded by 40 records.

In many cases the main problem is that the former data provider of Mensmeu 1.0 is different from the current EJP SOIL partner, thus the range of available data considered important/detailed/reliable is different also, which would have required a complete rewrite rather than a renewal, which most

[56] www.borisdaten.at
[57] https://dov.vlaanderen.be/
[58] https://www.thuenen.de/de/wo/arbeitsbereiche/bodenschutz-und-waldzustand/bodenzustandserhebung/
[59] http://dosoremi.hu
[60] https://www.geoportal.lt/geoportal/web/en
[61] https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Facilities-Tools/Dutch-Soil-Information-System-SIS.htm ; https://basisregistratieondergrond.nl/english
[62] https://www.vupop.sk/eng/index.php ; http://www.podnemapy.sk/default.aspx
[63] www.nabodat.ch
[64] http://soilmaps.it
[65] Piemonte: https://www.regione.piemonte.it/web/temi/agricoltura/agroambiente-meteo-suoli/geoportale-piemonte-carte-dei-suoli-dei-paesaggi-atlante-dei-terreni
Veneto: https://www.arpa.veneto.it/temi-ambientali/suolo/conoscenza-dei-suoli
Emilia-Romagna: https://geo.regione.emilia-romagna.it/cartpedo/
Toscana: http://sit.lamma.rete.toscana.it/websuoli/
[66] https://geoportal.vumop.cz/
[67] https://webapps.gissol.fr/geosol/

partners did not undertake. It is therefore not possible at present to validate the Mensmeu results based on the responses of the EJP SOIL WP6 partners.

Belgium, France, Germany, Hungary, Latvia, Lithuania, the Netherlands, Poland, Slovakia, Switzerland and the UK (Northern Ireland) declared to have a **soil monitoring system**. Belgium, in 2018, has declared having a soil monitoring system in establishment to monitor: Soil erosion, Loss of soil organic matter, Soil sealing, general properties. Denmark, in 2020, has declared having an established soil monitoring system to monitor: Loss of soil organic matter, Nutrient imbalance, general properties, plant nutrients. Germany, in 2018, has declared to have an established soil monitoring system in forest soils (Germany1) to monitor: Loss of soil organic matter, Nutrient imbalance, Soil acidification, Loss of biodiversity, Soil pollution / contamination, general properties, plant nutrients, soil pollution/contamination. Germany, in 2018, has further declared having a soil monitoring systemin establishment (Germany2), to monitor: Loss of soil organic matter, general properties, plant nutrients, soils salinity/alkalinity. Hungary, in 2020, has declared to have an established soil monitoring system to monitor: Loss of soil organic matter, Soil acidification, Soil pollution / contamination, Salinization / sodification, general properties, plant nutrients, soils salinity/alkalinity, soil pollution/contamination. Latvia, in 2020, has declared to have an established soil monitoring system to monitor: Organic carbon/organic matter, pH, general properties, plant nutrients. Lithuania, in 2020, has declared to have an established soil monitoring system to monitor: Soil erosion, Nutrient imbalance, general properties, plant nutrients. Netherlands, in 2021, has declared to have an established soil monitoring system to monitor: soil properties, functions and threats are derived afterwards. Poland, in 2020, has declared to have an established soil monitoring system to monitor: Loss of soil organic matter, Nutrient imbalance, Soil acidification, Soil pollution / contamination, salinization / general properties of agricultural, plant nutrients availability. Slovakia, in 2020, has declared to have an established soil monitoring system to monitor: Soil erosion, Loss of soil organic matter, Nutrient imbalance, Soil acidification, Soil pollution / contamination, Salinization / sodification, Soil compaction, soils used for energetic crops, peatland, general properties, plant nutrients, soils salinity/alkalinity, soil pollution/contamination. Switzerland, in 2021, has declared to have an established soil monitoring system to monitor: Loss of soil organic matter, Soil acidification, Soil pollution / contamination. UK (Northern Ireland), in 2020, has declared to have an established soil monitoring system to monitor: Loss of soil organic matter, Nutrient imbalance, Soil acidification, Soil compaction, general properties, plant nutrients, soils salinity/alkalinity, soil pollution/contamination. Recently, France, in 2021, declared having a soil monitoring system established in 2000 (France2), to monitor for all kind of land: loss of soil organic matter, soil contamination, soil biodiversity, general properties, plant nutrients and soil management practices. Austria and Norway have declared to have a monitoring system, intended as map contents upgrading. The Netherlands is starting a soil monitoring system for soil properties, the baseline survey has been performed in 2018 after an initial survey in 1998 (van Tol-Leenders, 2019).

The following countries declared that they do not have a national soil monitoring system: Finland, Italy, and Turkey.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Table 2.9 Summary of answers given in the CountrySIS survey on the Digital Database of Soil Properties

Country	Map	Point data	Accessibility	Data on	DB format	Metadata on	Metadata standards
Austria1	soil map of Austria (1:25.000)	with 12 reference profiles	open access	general properties, and salinity	PostgreSQL	Time (date) of soil survey / soil analysis, Data source	
Austria2		Point data	data available with restrictions	general properties, plant nutrients, salinity, and pollutants/contaminants	ORACLE	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	BORIS / Data key soil science, INSPIRE
Belgium	polygons	point-data (pit and bore)	open access with exception of contaminant and soil nutrients	general properties, plant nutrients, salinity	PostgreSQL	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	INSPIRE
Denmark	maps	soil profiles	not open access	general properties, plant nutrients, and pollutants/contaminants	MS Access, additional data as Excel spreadsheets	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis	-
Finland		soil profiles	not open access	general properties, plant nutrients, salinity, and pollutants/contaminants	csv	None	
France 1	Soil map of France (1:250 000 and 1:1M) (IGCS)	Point data/soil profiles/ polygons	Several owners with different rules of access	general properties, plant nutrients	PostgreSQL	Measurement units, time (date)/ soil analysis, land use	
France 2	Soil monitoring network (RMQS)	Point data/soil profile	Available with license	General properties, plant nutrients, contaminants, microbial diversity.	PostgresSQL	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, soil management information, site description	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

France 3	Soil test database (BDAT)	Point data	Available with license	General properties, and plant nutrients	PostgresSQL	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of analysis	
Germany1	forest soil maps	soil profiles	open access	general properties, plant nutrients, and pollutants/contaminants	PostgresSQL	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source	
Germany2		inventory of SOC soil profiles	not open access (project ongoing IN 2018)	general properties, plant nutrients, and salinity	MySQL	None	none
Hungary1	polygons	points	not open access	general properties, and salinity	Microsoft SQL, shp	None	
Hungary2		soil profiles	available by request	general properties, plant nutrients, salinity, and pollutants/contaminants	FOXPRO	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis	
Hungary3		mixed sampling on soil parcels	not open access	general properties, plant nutrients, and pollutants/contaminants	MySQL	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	
Hungary4		soil profiles	not open access	general properties, and salinity	RData, csv	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, unit, description of the soil property, reference for measurement methods	Metadata of SoilGrids was considered for providing metadata of MARTHA.
Italy	polygons	reference profiles and soil derived profiles	open access only the data owned by CREA	general properties, plant nutrients, and salinity	MS ACCESS	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	Italian Standard
Latvia1	polygon		not open access	general properties, and plant nutrients	Microsoft SQL server	Measurement units, time (date) of soil survey / soil analysis, ownership, other: "in our database all	

						agrochemical research data is well structured - lots of metadata columns are described there	
Latvia2		soil monitoring program plot based	not open access	general properties, plant nutrients, salinity, and pollutants/contaminants	open document (ods), dbf / shp	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis	LVS / ISO
Lithuania	soil maps	soil profiles	open access	general properties, plant nutrients, and salinity	dbf, shp	Soil analysis methods, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	
Netherlands	1:50.000 national soil class map	Soil profile classification and description, detailed soil surveys, sampling	open access	general properties, plant available nutrients, salinity, and pollutants/contaminants	Oracle, xls	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	
Norway	polygons	points	open access	general properties, plant nutrients, and salinity	PostgreSQL/GIS	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	ISO-standards for the analysis in the lab
Poland	soil maps	soil profiles	only soil maps are open access	general properties, plant available nutrients, salinity, and pollutants/contaminants	shp/dbf, MySQL	Time (date) of soil survey / soil analysis	no specific standards
Slovakia	mapping units	soil profiles	open access only selected datasets	general properties, plant nutrients, salinity, and pollutants/contaminants	MS SQL Spatial, Oracle	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Measurement units	Datasets: ISO 19115; Services: ISO 19119
Switzerland	soil maps	soil profiles	a subset of the data	general properties, plant available nutrients, salinity, and pollutants/contaminants	Oracle, xls	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership, Data visibility / access restriction	As given by the national law of geoinformation and respective document

						parameters, as well as a number of data model hierarchical metainfo (such as info on the owner of the site; project that use this site etc.)	s. The metadata is described in the data model.
Turkey	db is under constr uction			general properties, plant nutrients, salinity, and pollutants/contamin ants	Excel	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownershi p	Turkish Standards Institution (TSE) Soil Survey Manual
UK (Northern Ireland)	soil maps	soil profiles	soil maps only are available with fees	general properties, plant nutrients, and pollutants/contamin ants	Access	None	

2.7 Limitations of the synthesis

In this WP6 stocktake, we obtained 25 filled in questionnaires. Respondents were partners with different research focuses, some of whom did not have sufficient knowledge or access to the detailed description of national databases expected in the questionnaire. The answers given in the questionnaires do not necessarily cover the entire overview of national datasets for each of the EJP SOIL partners.

The validity of the collected data for a given country depends on the respondents' access to the databases and the professional knowledge of the data infrastructure (not all partners are data owners; moreover, partners can be at different soil data user levels).

As the main purpose of the WP6 stocktake was to collect metadata on georeferenced national data sources, the result does not include the description of national research databases collected and managed independently.

The completed questionnaires are less representative for some countries because they do not contain national databases or their description is incomplete. Data is not always provided for the whole country, as the data of experimental sites are local. However, the laboratory background of the attributed measurement data can be assumed the same and as the laboratory practice typical of the country. Therefore, we considered it representative as a data source for describing the methodology.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

2.8 Conclusions

The main conclusions of the WP6 stocktake for data sources:

1. Basic soil property (such as soil organic carbon, pH, particle size) databases are available in each country, but some of the countries use different set of measurement methods. The report displays these sets on maps and figures. Harmonized mapping across Europe is hampered if a country uses a country-specific set of measurements. Even in the case of a uniformly developed mapping methodology (like GSM), sudden value changes may occur at the border (transboundary issues). Spectral data collection on soil properties is widespread in only about a third of the EJP SOIL partners.
2. In less than half of the countries data on soil threats are available for soil pollution, compaction, water erosion, and organic matter decline.
3. Many data on soil properties are freely available, but their spatio-temporal resolution is very varied, in most cases, uncertainty quantification is not assigned. It would be necessary to compare the national databases and the LUCAS parameters to explore the reasons for possible discrepancies. The launch of the European Soil Observatory in 2020 can accelerate this comparison activity.
4. There are only a few databases available on the subject of measures to control soil erosion, 2-6 country reported positively in these issues. Most of the partners reported, that access to national soil management databases is limited and regulated by the national ministries of agriculture.

The main conclusions deriving by the analysis of the CountrySIS responses are:

5. Basic data are stored in databases of very different formats, with very different data standards. Their accessibility is variable among countries, among different data owners inside the countries, and for different type of soil data. General soil properties and plant nutrients are always recorded. This is not the case for data on soil salinity, pollution, and contamination.
6. Not all the countries having soil databases, also have a SIS and/or a soil monitoring system. The maintenance of a SIS needs skilled staff, which is not always available. Some countries reported to have a SIS in 2018, which is no more accessible in 2020. There is a general complain about the lack of skilled staff, lack of financial resources, lack of time.
7. Other complains are about the lack of communication/coordination between organizations, which makes it difficult to organize and maintain a national soil database and connected SIS and soil monitoring system, and the lack of common standards needed to integrate different soil data sources.
8. Finally, some countries reported also the lack of specific legislation for the legal implementation of soil surveys and soil information system, and specifically for the soil data protection and data ownership.

Answering to these needs is the main aim of the WP6 of EJP SOIL, and the present deliverable constitutes a starting point.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

2.9 Acknowledgements

We would like to thank all the EJP SOIL partners for their contribution, filling in the questionnaires and reviewing the reported summary results:

Elmar Schmaltz (Austria), Katrien Oorts (Belgium Flanders), Florent Hawotte (Belgium Wallonia), Lenka Pavl (Czech Republic), Anders Bjørn Møller (Denmark), Kadri Allik (Estonia), Elina Nurmi (Finland), Antonio Bispo (France), Florian Schneider (Germany), Béla Pirkó (Hungary), Lilian O' Sullivan (Ireland), Maria Fantappiè (Italy), Raimonds Kasparinskis (Latvia), Dalia Feizienė (Lithuania), Kees Teuling and Thalisa Slier (the Netherlands), Henrik F. Mathiesen (Norway), Bożena Smreczak and Artur Łopatka (Poland), Maria da Conceição Gonçalves (Portugal), Rastislav Skalský (Slovakia), Borut Vrščaj (Slovenia), Marisa Tello (Spain), Johanna Wetterlind (Sweden), Gina Garland (Switzerland), Sevinc Madenoglu (Turkey), Alex Higgins (United Kingdom).

3 Harmonised procedures for creation of databases and sharing soil data

Authors: J.D. Bulens, L.M. de Sousa, G. L'Abate, C. Calzolari, C. Le Bas, L. Donovan, K. Oorts, R. Yahiaoui, F. Ungaro, C. Di Bene, F. De Natale, B. Parisse, L. Gardin, C. Lattelais, F.M. van Egmond, M. Fantappiè, M.C. Andrenelli, N.H. Batjes

This chapter describes the general principles for seamless data exchange and sharing across borders and / or organisations. In principle, they can be applied to all domains, but here we focus on the soil domain. Particularly, we discuss technical and semantic conditions needed to achieve the soil data sharing in Europe.

Although, the regulation and legal aspects on how shared data can be used and what it is allowed to do with them is a crucial point, it is not discussed in this chapter because it is the topic of a future deliverable D6.2.

As stated above, harmonisation procedures consist of two parts that are very different from each other. Technical harmonisation is domain-independent and it has no relationship with the content, while technical coordination is required to establish a link between technical formats wherewith the data are stored. In semantic coordination, agreement is reached on the content in the specific domain, the soil domain. What are the parts describing that domain? What are the associated terms and definitions and what relationships exist between them? Semantic harmonisation is supported by technical methods and tools to record the content and store the data.

In the following sections, we list past and current European or international initiatives or projects dealing with soil data sharing. Then follows the description of the semantic modelling process and review of the code lists as part of the harmonisation process. There, we zoom in on the harmonisation process to derive harmonised data within the domain through transfer functions. The final sections cover the work environment for storing data, followed by data capture and data publishing.

3.1 Relevant previous or ongoing soil related initiatives

Since the early initiatives of soil mapping, the challenge of exchanging soil data and information has been recognised as a topical issue when moving from a local approach to a national, regional or global one. Most important bottlenecks were differences in naming, defining, and structuring the soil data. Aligning data and agreeing on a consistent and unambiguous manner to use and exchange information is referred to as harmonisation leading to standards.

Soil data harmonisation includes soil survey (soil description, soil observations, classification, mapping), soil analysis and field measurements (sampling design, lab methods, quality assurance), exchange protocols for digital data (model, metadata, vocabulary), and their interpretation, (transformation and evaluation. During the past decades, several global and European initiatives were taken to streamline and improve the use of soil information by addressing the main soil-related issues

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

such as soil description, classification, mapping, lab analysis and information exchange (Baritz et al., 2014, 2017). In the following section, an overview of these initiatives or projects is presented.

3.1.1 Organisations

3.1.1.1 Food and Agriculture Organisation of the United Nations (FAO)

FAO[68] played a major role in the second half of the past century, developing a legend for the 1:5 M Soil Map of the World (FAO-UNESCO, 1974, FAO, 1988) which was actually used as a classification system. In cooperation with the International Union of Soil Science (IUSS), the legend developed towards a proper classification system, the World Reference Base for Soil Resources, now in its fourth edition (IUSS Working Group WRB, 2015). Other widely used classification systems include USDA Soil Taxonomy (Soil Survey Staff, 1999) and the French Référentiel Pédologique (Baize and Girard, 2009).

For soil description, FAO has a long history with its Guidelines for Soil Profile Description, now at its fourth edition (FAO, 2006). Other well tested guidelines exist, which are widely used as reference, e.g. the USDA Soil Survey Manual last edition: (Soil Science Division Staff, 2017) and the USDA Field Book for Describing and Sampling Soils (Soil Survey Staff, 2012). The ISO Technical Committee 190 (soil quality) is also addressing the issue (ISO ICS 13.080).

3.1.1.2 European Commission (EC)

Many initiatives related to soils are carried out by agencies or research centres of the EC[69] and within the EU-research programmes.

3.1.1.3 European Soil Data Centre (ESDAC)

The European Soil Data Centre (ESDAC)[70] was established at the European Commission's Joint Research Centre in Italy, in order to answer to the soil data and information needs of the European Commission and of the European Environment Agency (EEA). ESDAC serves as a contact point for DG Environment on soil information.

ESDAC stores and distributes a wide variety of information on European soils, including datasets, legacy maps and applications. The structure of ESDAC builds on more than twenty years of soil related activities, starting in the late nineties with the creation of the European Soil Bureau Network (ESBN) as a network of national soil science institutes. The harmonisation of soil data was the main challenge when the European Soil Database (ESDB) and the soil profile analytical database for Europe (SPADE) were launched. The SPADE harmonisation exercise is ongoing (Kristensen et al., 2019).

68 http://www.fao.org/home/en/
69 https://ec.europa.eu/info/index_en
70 https://esdac.jrc.ec.europa.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

3.1.1.4 European Environmental Agency (EEA)

Within the European Environment Information and Observation Network (EIONET)[71], the operative network of EEA[72], the National Reference Centres (NRC) collect data of environmental relevance. NRC soil[73] is a technical network of soil experts from all European countries and is aimed at contributing on soil related indicators for EEA reporting. The reporting of indicators is organised though EIONET data flows. Apart from National Focal Points and NRCs, EIONET has also seven European topic centres[74] (ETCs) one of which is on Urban, Land and Soil Systems.

3.1.1.5 ISRIC – World Soil Information

ISRIC – World Soil Information[75] was created in 1966 to serve the world as a knowledge base on soil information and specifically at the time for the "Soil Map of the World" by FAO, ISSS (now IUSS) and Unesco. Since 1989, ISRIC has been accredited as WDC-Soils to the International Science Council (ISC) World Data System (WDS). ISRIC contributes to standard setting and database development within the soils domain. Areas of attention include quality control, standardization, harmonisation and analyses of world soil data. ISRIC has been actively involved with the development of the Soil Map of the World, the Harmonised World Soil Database (HWSD), and the WRB classification system. Besides developing pedology-based approaches (e.g. WISE and SOTER) to soil mapping it is elaborating statistically-based, predictive mapping (SoilGrids) approaches with associated spatial data infrastructures. Standardization of analytical method descriptions is an important element of the WoSIS (World Soil Information Service) procedures. Much of the work is done in partnership. ISRIC is a key contributor to the development of the federated GLOSIS system within the Global Soil Partnership (GSP), with special attention for data-interoperability and ontologies.

3.1.2 Projects

A number of initiatives and/or issues at global or European level have required improved harmonisation and sharing of soil data involving research activity in projects like GS Soil or SIEUSOIL.

3.1.2.1 Assessment and strategic development of INSPIRE compliant Geodata-Services for European Soil Data (GS Soil)

In the frame of eContentplus programme, a multiannual EU programme to make digital content in Europe more accessible, usable and exploitable, GS Soil project focused on the exchange of soil information (2009-2012). To this aim, the project addressed all the aspects of data organisation, data harmonisation as well as semantic and technical interoperability. The focus was on data from national

[71] https://www.eea.europa.eu/about-us/countries-and-eionet
[72] https://www.eea.europa.eu/
[73] https://www.eionet.europa.eu/countries/national-reference-centres/nrc-on-soil
[74] https://www.eionet.europa.eu/etcs
[75] https://www.isric.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

or regional institutions. Several use cases were analysed, especially across borders between neighbouring countries. Data from soil maps and soil profiles were examined with a special attention for harmonisation. The crucial issue of intellectual property rights, including scope for sharing data, was also addressed. The project resulted in a GS Soil portal based on operational SoilML application schema and tools[76]. Unfortunately, this portal is no longer accessible. A series of Best Practice Guidelines were also produced for the harmonisation of soil related information, creating and maintaining metadata for soil databases (Klug and Bretz, 2012).

3.1.2.2 SINO-EU Soil Observatory for Intelligent Land Soil Use (SIEUSOIL)

More recently, the H2020 SIEUSOIL[77] project (2019-2021), is aimed at developing sustainable soil management practices, implementing and testing a shared China-EU Web Observatory platform that will provide Open Linked Data to monitor the status and threats of soil. It includes the development of an ontology for soil in synergy with GSP Pillar 5.

1.1.1.1 ENVRI-FAIR

An important milestone in the data sharing processes is the definition of the 'FAIR Guiding Principles for scientific data management and stewardship' (Wilkinson et al. 2016). These principles require that all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people. They were officially recognized by the G20 Leaders in Hangzou (China, 4-5 September 2016), are supported by several international organisations (CODATA, RDA), and are the basis for the European Open Science Cloud development (EOSC)[78]. The purpose of EOSC is *to federate existing research data infrastructures in Europe and realise a web of FAIR data and related services for science, making research data interoperable and machine readable* following the FAIR guiding principles[79]. In this context, the H2020 ENVRI-FAIR project[80] focuses on the ESFRI Cluster of Environmental Research Infrastructures (ENVRI), which cover the main four subdomains of the Earth system (Atmosphere, Marine, Soil, Earth, Biodiversity/Terrestrial Ecosystems), with the overarching goal to connect it to the EOSC. In continuity with previous projects involving the ENVRI community[81], common data standards and policies for data life cycle, cataloguing, curation, provenance and service provision are being further developed[82].

---

76 https://inspire.ec.europa.eu/events/conferences/inspire_2011/presentations/workshops/8/GSSOIL_3_appl_schema_Eberhardt.pdf

77 https://www.sieusoil.eu/

78 https://eosc-portal.eu/

79 https://www.nature.com/articles/sdata201618

80 https://envri.eu/home-envri-fair/

81 https://envri.eu/current-envri-projects/past-envri-projects/

82 https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967

3.1.3 Networks and partnerships

3.1.3.1 International Union of Soil Sciences (IUSS)

A IUSS[83] working group on World Reference Base for soil (WRB) is active, maintaining and updating the most used international soil classification, which is now at the third edition, published in 2014. The WRB[84] is the international standard for soil classification system endorsed by the International Union of Soil Sciences. It was developed by an international collaboration coordinated by the IUSS Working Group. It replaced the FAO/UNESCO Legend for the Soil Map of the World[85] as international standard. The WRB borrows heavily from modern soil classification concepts, including Soil Taxonomy[86], the legend for the FAO Soil Map of the World 1988[87], the Référentiel Pédologique and Russian concepts. As far as possible, diagnostic criteria match those of existing systems, so that correlation with national and previous international systems is as straightforward as possible.

Since 2010, a IUSS[88] working group has been developing a ”Universal Soil Classification System (USCS)“ aimed at integrating the most used systems in the world (Micheli et al. 2016)**.** Another IUSS working group, on Soil Information Standards (SIS), has been activated in 2010 and is currently not very active due to soil data exchange activities in other groups/initiatives.

3.1.3.2 Global Soil Partnership (GSP)

At global level, the GSP (see paragraph 1.2.7) is working on the implementation of a Global Soil Information System (GLOSIS)[89] in Pillar 4 and 5, aimed to be a federated soil information system connecting national and regional soil information systems (FAO, ISRIC in press). GLOSIS[90] is composed by four main building blocks: a domain model, data exchange standards, nodes and support nodes and a “discovery hub”, i.e. a web-based gateway for data browsing. Conceptually, this model is closely related to INSPIRE and will share, as much as possible, the same international codelists. GLOSIS builds on and collaborates with the existing experiences of previous and ongoing programmes/ initiatives, e.g. SOTER, eSOTER, WoSIS, GlobalSoilMap, SIEUSOIL.

In the GSP Pillar 5 implementation plan[91] the creation of a Global Soil Laboratory Network (GLOSOLAN) has been introduced, in order “to build and strengthen the capacity of laboratories in soil analysis and to respond to the need for harmonizing soil analytical data”. Following the bottom up GSP approach, GLOSOLAN[92] is a participative network of laboratories sharing their knowledge and expertise for

---

[83] https://www.iuss.org/int-year-of-soils-2015/working-groups-for-iys/working-group-world-reference-base-for-soil-resources-wrb/ ; https://www.boku.wzw.tum.de/index.php?id=wrb&L=0 ; https://sites.google.com/a/vt.edu/iuss1-4_soil_classification/home
[84] http://www.fao.org/soils-portal/data-hub/soil-classification/world-reference-base/en/
[85] http://www.fao.org/soils-portal/data-hub/soil-classification/fao-legend/en/
[86] http://www.fao.org/soils-portal/data-hub/soil-classification/usda-soil-taxonomy/en/
[87] http://www.fao.org/soils-portal/data-hub/soil-classification/fao-legend/en/
[88] https://www.iuss.org/int-year-of-soils-2015/working-groups-for-iys/universal-soil-classification-system-working-group/ .
[89] http://www.fao.org/global-soil-partnership/areas-of-work/soil-information-and-data/en/
[90] https://glosis.org/
[91] http://www.fao.org/3/a-bs756e.pdf
[92] http://www.fao.org/global-soil-partnership/glosolan/en/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

developing harmonised standards. The network is organised in regional networks (RESOLAN) based on national networks. At European (Eurasian) level, the European and Eurasian Soil Laboratory Network (EUROSOLAN) is active since 2019. Wet chemistry and dry chemistry (soil spectroscopy) are addressed and harmonised Standard Operating Procedures (SOP) are being produced yearly.

3.1.3.3 InteroperAble Descriptions of Observable Property Terminology Working Group (I-ADOPT WG)

In the framework of the Research Data Alliance (RDA), which is dedicated to building social and technical bridges to enable the open sharing and re-use of data, several working groups or related organisations are active. One of which is the I-ADOPT[93] working group which focuses on creating a community-agreed framework for representing observable properties by bringing together groups that have been working on developing terminologies to accurately encode what was measured, observed, derived, or computed. Another is the Agricultural Data Interest Group (IGAD)[94] who have several working groups working on standardising vocabularies on different agricultural topics. One of the topics discussed has been soil.

3.1.3.4 International Soil Moisture Network

Thanks to the combined effort of Global Energy and Water Exchanges Project (GEWEX[95]), the Committee on Earth Observation Satellites (CEOS[96]), the Global Climate Observing System - Terrestrial Observation Panel for Climate (GCOS-TOPC[97]), the Group of Earth Observation (GEO[98]), and the Global Terrestrial Network on Hydrology (GTN-H[99]). The International Soil Moisture Network[100] is an international cooperation to establish and maintain a global in-situ soil moisture database. This database is an essential means for validating and improving global satellite products, and land surface, climate, and hydrological models.

3.1.3.5 EU-directive INSPIRE

An important initiative on data sharing is the INSPIRE Directive which came into force in 2007 in the European Union. The European Commission and the European Environment Agency noted the difficulty of obtaining data on the environment which were compatible between them and between the Member States. Thus, the objective of the INSPIRE Directive was to develop a European infrastructure for spatial information needed for environmental policies. This information was divided in themes listed in the three annexes of the Directive. Soils are part of the Annex 3. The main principle

[93] https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg
[94] https://www.rd-alliance.org/groups/agriculture-data-interest-group-igad.html
[95] http://www.gewex.org/
[96] http://www.ceos.org/
[97] https://public.wmo.int/en/programmes/global-climate-observing-system/terrestrial-observation-panel-climate
[98] http://earthobservations.org/
[99] http://www.gtn-h.info/
[100] https://ismn.earth/en/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

of the Directive is to develop a network of spatial data infrastructures to find, share and exchange environmental data from local, to national and European levels, which has to follow technical guidelines to ensure interoperability. Interoperability in INSPIRE means the possibility to combine spatial data and services from different sources across the European Union in a consistent way without involving specific efforts from humans or machines. Therefore, INSPIRE data specifications are defined for soil both in a natural and in a conceptual scheme language (TWG-Soil, 2013). From 2010 to 2013, a thematic working group on soil was mandated by the European Commission to elaborate the data specifications for soil data. It was composed of several experts from different member states with an expertise on soil data, geomatics, and data modelling. Their work resulted in the publication of soil data specifications in 2013 which comprises the definition of a soil data model for INSPIRE.

The data specifications were based on a common template used for all data specifications, which has been harmonized. To enhance semantic interoperability, the use of the soil classification scheme WRB (World Reference Base for Soil Resources) and the FAO horizon notation scheme were proposed as primary classification systems, whereas the use of other currently used (local, regional, national) classification systems is also provided for. Publishing the data according to the data specifications, interoperable, can be achieved by either changing (harmonizing) and storing existing data sets or mapping and/or transforming them for publication in services in the INSPIRE infrastructure. It is important to note that 'interoperability' is understood as finding the data and providing access in a standardised manner to spatial data sets through network services, typically via Internet. Finding is secured by creating and providing metadata for these data. Metadata provide not only what the data are about, but also where to access to them (the data repository) and the license that is used to share the data defining the conditions and restrictions of use. Metadata are exposed in discovery services, for INSPIRE this is the INSPIRE geoportal. Metadata can be harvested by other discovery services that also use the ISO metadata standard or that can map to this standard. Note that this does not mean it can be found by big search engines like Google, Bing, etc. since they are not designed to search for data within data repositories.

An important notion is that INSPIRE specifies what and how to structure, exchange and share data within an infrastructure, but does not contain data itself, that is done by the data providers and/or owners, regional, national and/or international. It is a legal framework implemented by the EU member states. On a technical point of view, the main objective of the Directive is to be based on existing international standards such as those of the Open Geospatial Consortium (OGC)[101] or the International Standards Organisation (ISO)[102].

It is expected that by using the INSPIRE infrastructure, the users can use the soil data standardised and will spend less time and efforts on understanding and integrating data when they build their applications based on data delivered in accordance with INSPIRE.

[101] https://www.ogc.org/

[102] https://www.iso.org/home.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

3.1.3.6 International Standardisation Organisation (ISO)

The ISO standards[103] cover most of the classical procedures for soil chemical, physical and biological properties. However, these are not always followed or they are interpreted differently, due to different reasons. Within the frame of the GSP, Pillar 5 provides "the mechanisms for developing and exchanging globally consistent and comparable harmonised soil information". This refers to soil profile description, analytical data, and derived products. Soil data exchange is addressed by ISO (ISO 28258:2013) in a general manner and by other initiatives.

3.1.3.7 Open Geospatial Consortium (OGC)

The OGC Soil Data Interoperability Experiment (SoilIE), conducted under the auspices of the OGC Agriculture Domain Working Group in 2015, had the objective of developing and testing a soil standard that harmonised existing standards defined in Europe and Oceania. This IE evaluated existing models and proposed a common core model, including a GML/XML schema, which was tested through the deployment of OGC web services and demonstration clients. Time constraints and limited participant resources precluded extensive modelling activities in the IE. During the SoilIE, participants from Europe, North America and Oceania mapped data in their soil databases to the SoilIE XML schema. Multiple OGC Web Feature Services (WFS) delivering soil observation data using the XML schema were established, along with OGC Web Processing Services to allow on-line derivation of new data. A Soil IE Demonstrator[104] using the services was developed by the Centre for eResearch and Digital Innovation (CeRDI) - Federation University Australia.

3.2 Conceptual and data model, ontologies, general principles

3.2.1 The principles of conceptual modelling

For capturing real world phenomena, it is since long common practice to create a replica in a working environment to better understand, to have a means of communication or possibility to improve a design. Today, in the digital era, this environment is a digital environment, so capturing data is basically creating a digital copy of what is known of that environment. But how to make that digital copy that describes the real world? Just capturing data is not enough. We need to model it providing an abstraction of the real world that is fit for purpose. That abstraction starts with describing the concepts (conceptual modelling), describing the relationships (together referred to as ontologies) and implementing it in a digital environment (data modelling). This should be done in a way that is both human and machine readable. Human readable because when 'soil' is described in a structured way, it helps to communicate about it and to better understand the soil data. Machine readable is necessary to use soil data in algorithms or to build applications that use them. This chapter aims to guide through this process that allows not only to capture data about soil in a structured way, but also to use it in practice.
The typical process is to first conceptualize into an abstract model. Then to translate it into a technical

[103] https://www.iso.org/technical-committees.html
[104] http://data.cerdi.edu.au/soil_demo.php

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

model that can be implemented in a technical working environment. The technical model is typically reflecting the structure that is used to exchange the data.

THE FEATURE OF INTEREST

A feature of interest refers to a real object or phenomenon. In modelling there are many ways to name the 'feature of interest' mostly depending on personal preferences or on what 'technical' language you use. The most commonly used are 'Object', 'Class', 'Category' and 'Entity' that can be defined as an abstraction of a 'real world thing'. They are used all in basically the same meaning as a reference to the feature of interest. An important notion is that the feature of interest is not 'data' but information, in principle a set of data that is related and already has meaning.

In this document the naming of the feature of interest can be one of these, but they are in principle interchangeable.

Extracting those phenomena from the real world starts with creating abstractions of the components or objects that make up that phenomenon. These are then described in a conceptual model containing the 'features of interest' (objects/classes/categories/entities) with properties and the mutual relationships within them, so that they fit into the context of the intended use. To avoid that conceptual models only exist in the minds of people and are communicated verbally and often inaccurately in this process, these are written down and stored for wider dissemination[105]. In the ISO standards for geographic information, conceptualisation starts with identifying the Universe of Discourse: a selected piece of the real world that a human being wishes to describe in a model to be fit for purpose.

The selected 'pieces of information' to be fit for purpose are identified as the 'smallest' objects of information relevant within the so-called Universe of Discourse. The universe of discourse may include not only objects such as watercourses, lakes, islands, property boundaries, property owners and exploitation areas, but also their attributes, their functions and the relationships that exist among such features. The figure below describes the relationship between modelling the real world and the resulting conceptual schema.[106]

[105] Ewa Dębińska and Piotr Cichociński 2007; Conceptual Modelling of Real Estates for the Purposes of Mass Appraisal; in fig proceedings 2007 (https://www.fig.net/resources/proceedings/fig_proceedings/fig2007/papers/ts_4d/ts04d_01_debinska_cichocinski_1290.pdf)

[106] INSPIRE data specifications D2.5 Generic conceptual model; 2013; Drafting team data specifications (https://inspire.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3_vs_3.4rc2.pdf)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

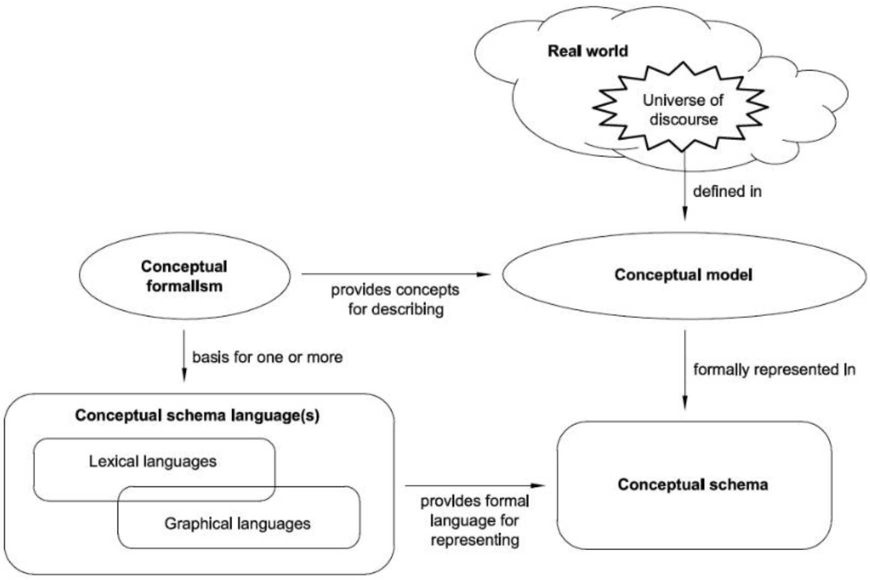


Figure 3.1 Conceptual modelling the real world (ISO19101)

The figure provides a graphical illustration of the role of conceptual modelling in representing geographic information by describing the relationship between modelling the real world and the resulting conceptual schema.

A universe of discourse is a selected piece of the real world that a human being or a community wishes to describe in a model. The universe of discourse may include not only spatial objects such as watercourses, lakes, islands, property boundaries, property owners and exploitation areas, but also their attributes, their operations and the relationships that exist among such spatial objects. A universe of discourse is described in a conceptual model.

The conceptual schema is a rigorous description of a conceptual model for some universe of discourse. A conceptual schema language is used to describe a conceptual schema. A conceptual schema language is a formal language parsable by a computer or a human being that contains all linguistic constructs necessary to formulate a conceptual schema and to manipulate its content. A conceptual schema that defines how a universe of discourse shall be described as data and operations is called an application schema.

A conceptual schema language is based upon a conceptual formalism. The conceptual formalism provides the rules, constraints, functions, processes and other elements that make up a conceptual schema language. These elements are used to create conceptual schemas that describe a given information system or information technology standard. A conceptual formalism provides a basis for the formal definition of all knowledge considered relevant to an information technology application. More than one conceptual schema language, either lexical or graphical, can adhere to and be mapped to the same conceptual formalism. In INSPIRE, every conceptual schema will be modelled using the conceptual schema language UML.

A conceptual schema is independent of physical implementation technologies and platforms. The conceptual model is a consolidated base that as a base won't change that much, but can be extended when additional context within the Universe of Discourse becomes relevant. To describe this

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

conceptual model, which also is referred to as an information model or also as a semantic model, currently there are two main schema techniques that are used to visualise it. One, more traditional, is the class diagrams as used within the Unified modelling Language (UML) referred to as meta-modelling in a Model-Driven Architecture (MDA). The other is the development into a more web based technique using ontology or knowledge graphs (KG) that depict the triplets 'subject-predicate-object' used for ontologies. The next section is explaining more on that subject with an overview of IT-developments in the past decades. Both schema techniques show the relationships between objects and their properties and objects to objects within the Universe of Discourse.

The next step, to produce a technical model, is the first step for implementation. How this looks like can be manifold, also depending on how you want to use the data. Important is that whatever implementation you select, it should be a proper translation of the conceptual model that is used as a consolidated base for all implementations. This is shown in Figure 3.2.

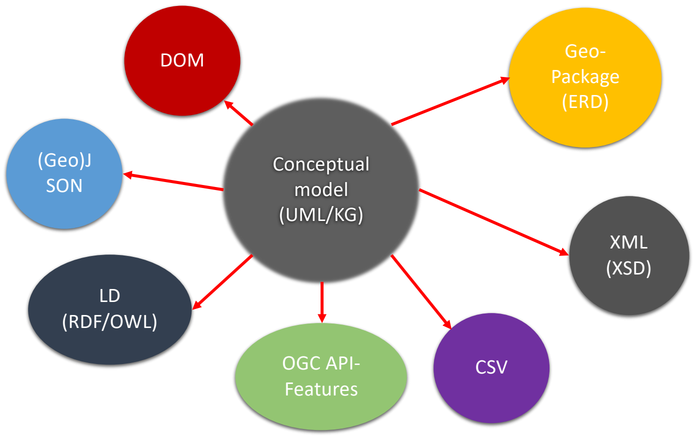


Figure 3.2 The conceptual model as a base for several implementations

The schema in Figure 3.2 shows different implementation techniques, but keep in mind that they can differ a lot in its expressiveness to carry the underlying information. There is not a single 'best' choice. The best choice is primarily determined by the intended use and selected working environment. The implementations are created to exchange data, so either sending or receiving data. An important notion here is that an exchange format is to transfer data from one to another location. Depending on the application that processes the data, you can use the data directly, but you can also save it in your own environment first. When saving, this does not necessarily have to be in the same structure as it is exchanged. In some cases, it can be better to use your own structured database in your own work environment, but of course without losing the necessary information carried by these data.

3.2.2 IT developments towards ontologies in information science, an overview

Industries understood the power of computers to process data soon after World War II. Business oriented hardware and software proliferated throughout the 1950s, at first without coordination between vendors. In 1959, several companies in the United States assembled a consortium named CODASYL (Conference on Data Systems Languages) with the purpose of defining a programming language for data systems that could be executed on multiple hardware platforms, i.e. independent

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

from vendors. One of the results was the COBOL (COmmon Business-Oriented Language) programming language. If at first it had little clout with the software industry, once it was made mandatory by the United States government (Ensmenger 2009) it became a de facto standard.

Data stored by computers those days amounted to little more than collections of files, each storing a set of records. Soon enough industrial and governmental information systems grew in complexity past such simple structures. As the 1970s dawned, various attempts emerged towards more abstract and complex ways to represent data stored by computers. Eventually, Chen (1975) proposed the Entity-Relationship (ER) meta-language, that finally broke in as a popular choice. Entity is something that exists, a being or a particular unit, relationship is a connection or association. ER is a graphical language, providing constructs to express categories of data, their attributes and the relationships between categories. While simple, ER completely abstracts the description of data from the underlying software or hardware.

Chen's choice of words was not at hazard. In 1970, Codd had introduced the concept of "relational database", defining rules for data management software that went beyond earlier file-based systems. The first implementation of Codd's vision was released in 1976 by IBM, the Multics Relational Data Store. In 1979, a small company named Relational Software released Oracle, which grew enough to even take over the name of the company. ER and relational databases proved a perfect match, providing the theoretical and practical facets to data management. Together they swept the software industry and computer science *curricula*.

In 1967, researchers at the Norwegian Computing Centre introduced a language for computer simulation -- Simula -- that included the concept of objects, classes of objects and class inheritance (these concepts are explored in more detail in the following section). Simula was not a success with the industry, but proved immensely influential on subsequent programming languages. In 1980, Smalltalk was released, product of an effort at Xerox towards an educational programming language. Smalltalk not only adopted the concept of objects from Simula, it made them its central paradigm. By the middle of the 1980s the introduction of industry grade languages like C++ and Eiffel made object-oriented programming a staple of software development.

At the dawn of the 1990s a more fundamental understanding of software development came about. First Powers (1991) and then Gruber (1993) proposed the direct application of Ontology to computer science. As a branch of Metaphysics developed in classical Greece, Ontology studies existence and being. It contrasts concepts such as particular versus universals, abstract versus concrete, properties versus relations, states versus events. The term ontology became first popular within the artificial intelligence community and later in computer science to signify an abstract representation of real-world concepts pertaining to a particular domain or field. It can be seen as the grammar that describes the entities and its relations in a domain or data model (described in more detail in 3.3).

The rapid growth of object-oriented programming fuelled the demand for novel abstract means to develop and document software. Rumbaugh et al. (1991) and Booch (1993) proposed the earliest infrastructures towards this end. Reunited under the Rational Software Corp. these and other researchers would develop such concepts into the Unified Modelling Language (UML). UML matched object-oriented programming just as ER had matched relational databases two decades earlier. But UML is a far more powerful and extensive language, allowing the abstraction of a wide range of constructs, such as class inheritance and composition, all with an expressive graphical meta-language.

UML largely provided the infrastructure for applied philosophy envisioned by Powers and Gruber. UML was adopted as a standard by the Object Management Group (OMG) in 1997, at a time when it already featured at large in computer science curricula.

At the dawn of the 21st century, the UML standard was pushed into an even higher level of abstraction. In 2003, the IEEE Software journal published a series of articles advocating a novel software development method named Model-Driven Development (MDD) in which domain models are the primary products, and source code, the code used to create it, is a by-product (Atkinson and Kühne 2003, Selic 2003). This idea was not entirely new, as various companies had since the 1980s proposed software to generate source code from graphical models (commonly known as CASE tools). What MDD brought anew was the extension of UML into meta-modelling, using abstractions such as categories of categories to capture the essential aspects of a knowledge domain. A broader discipline covering MDD, CASE tools and more became known as Model-Driven Engineering (MDE) (da Silva 2015).

In 2005, UML version 2.0 was released, including an entire infrastructure (primitives and methods) dedicated to meta-modelling named Model-Driven Architecture (MDA). With MDA, the core UML primitives can be specialised through a special primitive: the stereotype. A semantically related set of stereotypes can be gathered into a UML Profile, thus constituting a domain-specific lexicon, i.e. an ontology. MDA was almost immediately adopted by the industry in general and has since been used by various institutions to issue standards. Noteworthy are those issued by the Open Geospatial Consortium (OGC), many of which were also adopted by ISO. The INSPIRE model[107] is also specified with the MDA infrastructure.

In parallel to the efforts of the OMG, the World Wide Web Consortium (W3C) also worked towards an ontology infrastructure. The W3C was primarily concerned with the exchange and automatic processing of data in the age of the internet. It started by specifying the Resource Description Framework (RDF), a data encoding mechanism inspired on Knowledge Engineering principles. In RDF, each datum is a triplet mimicking natural speech: a subject, a predicate and an object. RDF was underpinned by a meta-language named RDF Schema, that specifies the set of constructs used to encode data as RDF triples. It encompasses ontological concepts such as category (class), property (domain, range, etc) or inheritance (sub-class).

When the first full RDF specification was released in 2004, the W3C had already started working on a more abstract infrastructure for meta-modelling. With a purposeful name, Web Ontology Language, and a catchy acronym, OWL, it presented a novel approach to ontology modelling. OWL is less abstract than UML, resulting from a process focused on the practical aspects of data exchange. An ontology specified with OWL is itself a RDF document, thus it is also a resource. It can easily be exchanged over the web and used together with other ontologies. The collection of standards issued by the W3C around RDF and OWL became known as the Semantic Web.

More recently the concept of Linked Data (Berners-Lee, 2006) gained popularity. It is not a standard by itself, rather a set of principles for data exchange based on RDF. In a nutshell, Linked Data prescribes the adoption of the Semantic Web standards with the strict use of Uniform Resource Identifiers (URIs) as identifiers for all resources. Each element of each triple and each ontology construct are thus

[107] https://inspire.ec.europa.eu/portfolio/data-models

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

universally and unequivocally identifiable, and more importantly, referenceable through the HTTP protocol.

Linked Data and the Semantic Web never reached the ubiquity of UML and MDA in the industry. The Semantic Web standards are primarily based on text, they are less expressive than the graphical grammars and not focused on software development. Therefore, fewer support tools exist. However, the semantic web has become popular in academia, both in computer science as in the emerging discipline of data science.

3.2.3 Core Ontology abstractions

Ontology is a millenary philosophy discipline with various twists and turns through history and a myriad of unresolved dissensions. It is therefore important to realise that the concepts absorbed in information science are not universally accepted within the Ontology discipline itself. However, they enclose the metaphysical principles supporting the development and modelling of information ontologies.

A core idea of Ontology is the contrast between universals and particulars (Honderich, 2005a). A universal is a category of entities that can be exemplified by various particulars. A particular is an entity that can usually be sited at a particular time and point in space. Universals are therefore more conceptual (or metaphysical) and particulars more physical. For instance, if "fruit" is a universal, "apple", "orange" and "pear" are particulars.

In information science (see figure 3.4), universals mostly appear with the name Class, a term common to both UML and OWL (in the old ER meta-model these where the Entities). A Class is a category of entities that share a common set of characteristics. In MDA and object-oriented programming, particulars are known as "instances", "class instances" or "objects". In the semantic web, particulars are called "individuals", a term that is also found in Ontology. The concept of Class in UML is somewhat broader since it can also specify behaviours that are common to its instances. However, this aspect is more relevant to programming than information science.

A further core concept in Ontology is that of "property" (Orilia and Paolini Paoletti, 2020) which is employed in similar sense in information science. A property conveys a specific characteristic of its bearer, expressing what the bearer is like. In information science, both universals and particulars have properties. At the universal level, they express a type of feature, whereas for the particular they assign a concrete value to that feature. For instance, if the "fruit" universal has the "colour" property, the particulars instantiate it with "red" for "apple", "orange" for "orange" or "green" for "pear". Properties in the semantic web and MDA have a determined type, usually an atomic computer system type (i.e. floating point, string), or a combination of these.

Another Ontology concept taken literally in information science is that of “relation” (Honderich, 2005b). Relations express how different entities stand to each other. This term gave the name to the Entity-Relationship meta-model but is often referred as "association", particularly in UML. In information science, relations have the critical facet of "cardinality", already present at the time of ER. With cardinalities, information ontologies express how many particulars of a certain universal can relate to one particular of another universal. For instance, if "fruit" and "seed" are universals, a relationship specifies that a particular of "fruit" can have (i.e. relate to) many particulars of "seed" but a particular of "seed" relates to a single "fruit" particular. Cardinalities are essential to structure storage and validate data in computer systems.

Other Ontology concepts were absorbed in a less straightforward fashion in information science. Most relevant among these is ontological dependence (Tahko and Lowe, 2020), stating that certain entities cannot exist without the existence of another (usually related) entity. Ontological dependence is sub-divided into sub-concepts: rigid dependence, that refers to a specific particular and generic dependence, referring to a category of particulars (or universal). In information science rigid dependence is usually expressed through cardinalities in relations. However, UML provides a specific construct akin to rigid dependence named "composition".

Generic dependence appears in information science in the form of class hierarchies. It features prominently in OWL and UML, signifying that a child class yields all the same properties and behaviours of its parent class. Using again the same example, the universals "pomes" and "citrus" can yield generic dependence from the universal "fruit". In this case, "apple" and "pear" are particulars of "pomes", whereas "orange" is a particular of "citrus". This feature is referred by the names "inheritance" and "generalisation" in information science. The Ontology discipline also conceives generic dependence as a vehicle to hierarchic structuring, distinguishing between more fundamental entities and secondary ones.

Figure 3.3 presents the notations used in three different meta-languages for key Ontology concepts adopted in information science. The limited extent of ER is patent, even though it is arguably still the most popular today. The textual nature of OWL also contrasts with the other two (although graphical notations have been proposed, see ahead).

In terms of tool support, it is also relevant to note that ER remains far ahead, with a myriad of commercial and open-source tools available. ER's simplicity much contributes to this. For the UML language, two tools stand out: Enterprise Architect, a commercial offer, and Papyrus, an open-source plug-in for the popular Eclipse development environment. As for OWL the open source Protégé dominates the landscape with TopBraid a commercial alternative.

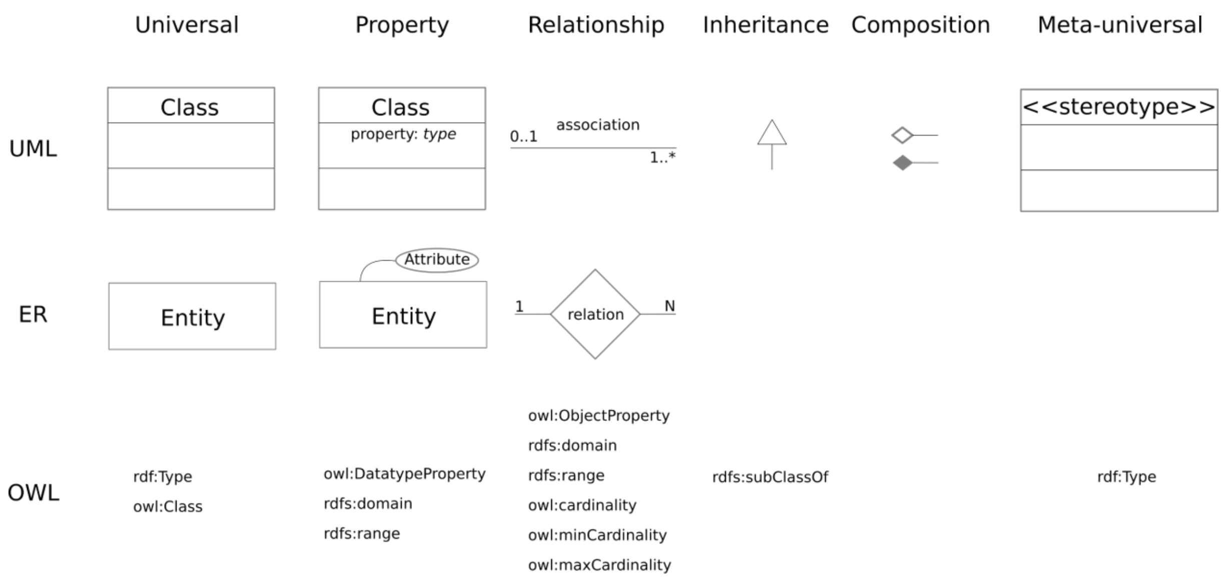


Figure 3.3 Notation of key ontological concepts in various information grammars.

3.2.4 The Modelling Process

The process of devising an ontology, applying the principles of conceptual modelling, customarily follows several different phases. They are rooted in the wider processes of software architecture and engineering (Bernard, 2012), going well beyond the concerns of information science. This section reviews three of these phases: (i) conceptual or domain modelling, (ii) technical or data modelling and (iii) data exchange or encoding.

The development of a domain model is the first phase, in which the discourse and concerns are most abstract and distant from technical aspects but need to include domain experts, in case of INSPIRE Soil these domain experts are the soil scientists. Ideally, no technical or implementation requirements are considered at this stage. In the software industry, this phase is supported by interviews and textual descriptions of the target business, its value creation processes and involved information entities. These assets account the human perception of the domain in natural language. Methodologies like Archimate (Lankhorst 2009) or ISO/IEC 42010 (ISO 2011) set out rules on how to formalise natural language into a domain model. They lay out the information entities relevant to the domain, their properties, relationships and hierarchies. Graphical meta-languages like UML and Archimate provide for a rich and expressive representation of a domain, which in theory can be a basis for discussion with stakeholders not versed in computer or information sciences. In the software industry, the domain model is also referred to as information architecture. For INSPIRE Soil a domain model was described in UML. Important to note is that each model is limited to the defined boundaries of its Universe of Discourse with the level of detail chosen for it. It means that, depending on its intended use, it can evolve into a more detailed and comprehensive domain model or ontology. Especially for INSPIRE, the domain model for soil is bound to the common level that is common for use within the European member states, which means that it needs to be extended for many more specific or national applications.

From a strict information science perspective, the result of the domain modelling phase is usually referred as an ontology. The goal is precisely the same: formalise and synthesise human knowledge in a particular domain using a standardised meta-language.

Information ontologies can also be represented with OWL, but are encoded as text documents. Therefore, it lacks the expressiveness of graphical languages, possibly being this the main reason why it is not as popular in the software industry. Notations for the graphical representation of OWL ontologies have been specified (Kendallet al. 2009, Heon 2020), but have not taken root. Another possibility is to develop the domain model with UML and later convert it into OWL, for instance applying the ISO standard 19150-2 (ISO 2015) as is done for GLOSIS.

The second phase concerns the design of data structures that can easily store data compliant with the domain model in a software system, as an implementation of the domain model. These structures must cope with the instances (i.e. individuals) of each information entity identified in the domain model. In the software industry, this phase traditionally equates to the translation of a UML model into a relational database model as one of the possible implementations. There have been various efforts by the software industry to create object-oriented databases, but none has gained popularity, as no concrete standard in this domain ever emerged. Thus, the need to fall back to a simpler formalism like a ER diagram or a collection of SQL statements. UML is considerably more sophisticated, therefore its translation into a relational model may not be straightforward. This is especially critical regarding class generalisation often forcing compromises into data structures that may not faithfully represent the domain model.

Interestingly, a parallel between information science and the practices of the software industry is not as clear-cut in the data modelling phase. Starting with the semantic web, an OWL ontology is itself a data model. Built on RDF primitives, an OWL ontology is ready to use directly in a data store. Moreover, OWL ontologies commonly include themselves the individuals that make up the thesaurus or code lists of the domain. More on this with respect to soil is described in chapter 3.3.

The final phase is possibly the most technical, concerning the method(s) of exchanging data compliant with the domain model. This phase is usually absent in the software industry, where the concern is primarily to support processes of value creation within an institution. Data exchange comes down to a specification on how to persistently encode datasets, making sure that they remain understandable by whomever retrieves it. The operation of persistently encoding data is also known as "serialisation" in computer science.

Various standards exist providing encoding grammars that can be used to automatically create a domain-specific format. The most common of such standards used in the MDA universe was actually specified by the W3C: the Extensible Markup Language (XML) (Bray et al. 1998). As with RDF, the W3C intended with XML to create a format readable by both humans and machines. The respective grammar, XML-Schema, was issued a few years later (Thompson et al. 2001) and can be used to map the features of a UML model (e.g. classes, attributes, relations) into the contents of a XML document. At the same time, a company called State Software specified a data format named JavaScript Object Notation (JSON) to facilitate communication between internet browsers and servers. JSON was almost immediately adopted by the internet software industry and became a de facto standard in 2013 (ECMA 2013). JSON yields similar features to XML, but is considerably more compact and lightweight.

Various specialisations of these data encoding standards are worth mentioning. The OGC specialised XML for the exchange of geo-spatial data with the Geographic Markup Language (GML) (OGC 2000). GML became a core feature of the web services standards specified thereafter by the OGC. The Web Features Service 2.0 standard introduced the concept of Application Schema, a further infrastructure for domain-specific data encoding that extends GML. The Internet Engineering Task Force (IETF) hosted an initiative that in 2007 produced the GeoJSON specification, a JSON specialisation for geo-spatial data (Butler et al. 2016).

As with the data modelling phase, the data exchange phase is also tacitly dispensed in the Semantic Web. Since all of its products are RDF triples, an OWL ontology is itself encoded in an exchangeable data format, as so any individuals expressed as triples. The W3C started by specifying and encoding based on XML, simply named the RDF/XML. But various other triple encoding formats emerged in the meantime, among which: Turtle (Beckett et al. 2014), a format favouring ease of read by humans, N-Triples (Beckett 2014) a parsing-friendly format and JSON-LD (Sporny et al. 2019), a specialisation of JSON.

3.3 Data standardisation - vocabulary and codelists

3.3.1 What about data?

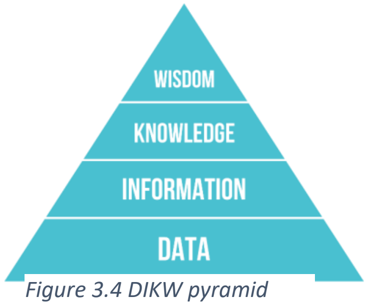


Figure 3.4 DIKW pyramid

We all know, or at least think we know, what we mean when we use the word data in our daily work. At the same time data have multiple facets that are important if they are to be converted into information. It starts with the notion that is in an easy way depicted in the DIKW pyramid (fig. 3.5). It provides the structural and / or functional relationships between data, information, knowledge and wisdom. "Information is generally defined in terms of data, knowledge in terms of information and wisdom in terms of knowledge"[108].

It shows that data is the foundation to build information on. Use of data cannot be seen without the connection to metadata.

Data are:

- the data themselves
  - observed or measured values
  - assigned or calculated values
- the metadata
  - the information about the data
  - the quality and uncertainty of the data (an aspect that is often not included).

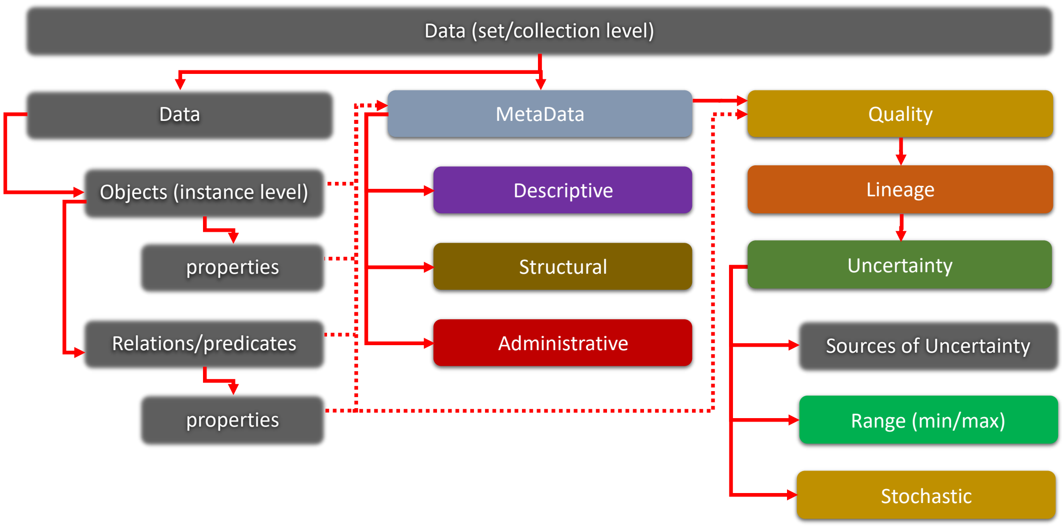


Figure 3.5 data, metadata and composing elements (Bulens et el., 2021)

108 https://en.wikipedia.org/wiki/DIKW_pyramid#cite_note-Rowley-1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

3.3.2 Data (themselves)

Basically, the data itself forms the primary material to work with. In general, the values are measured, observed, assigned or calculated and represent a characteristic of a property.

3.3.3 Metadata

Metadata is data about data. Metadata should provide you with all the information you have to know to work with the data. They consist of 'structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource, especially in a distributed network environment'[109]. Metadata is usually expressed by metadata elements and categorized in a number of different (overlapping) types:

- Essential metadata: method of measurement, units of measurement, measurement/observation time, postprocessing methods, precision, location, etc.
- Descriptive metadata, describing an information resource for identification and retrieval through elements such as title, author, and abstract.
- Structural metadata, documenting relationships within and among objects through elements such as links to other components (e.g., how pages are put together to form chapters). These include observation and measurement methods, units of measurement and other where relevant.
- Administrative metadata, for managing information resources through elements such as version number, archiving date, and other technical information for purposes of file management, rights management and preservation.

Metadata can be provided on the level of datasets, objects and properties, but all serve the same purpose. There are many standards to describe metadata elements. To mention the most important or influential ones:

1. the 'Dublin Core' for networked resources
2. for the spatial domain the 'ISO 19115:2003 Geographic information -- Metadata standard' defines how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use.

3.3.4 Quality & Uncertainty

Data are usually the result of direct measurements, observations and derived as outcomes of transformations or calculations. Due to measurement errors, structural errors in our models, and errors in the estimated model parameters, a user will always be confronted with uncertainty. In Ten Wehrens et al. (2021), more details are given on how to cope with uncertainty in a Digital Twin setting.

In short, the diagram above (fig. 3.6) shows all aspects of data that are to be considered when working with data in any context.

[109] https://en.wikipedia.org/wiki/Metadata

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

3.3.5 Observations and Measurements

Special attention must be paid when we want to use observed or measured data in an ontology. It is important to realise that except for the property/value (including the Units of Measure) itself the method used is not part of the domain ontology. Nevertheless, when using that value one needs to know how these values are obtained. All data proving information are regarded as metadata of the value on the 'instance' level. In that sense they do relate to the domain ontology as metadata and are usually included in the domain ontology. ISO 19156 is an international standard for structuring information on observations and measurements (Cox, 2013). It explicitly decouples procedures (methods), properties and results and helps keeping the number of properties defined in the model manageable (Ritchie et al., 2016). The Observations & Measurements (O&M) standard defines a domain-independent conceptual model for the representation of (spatiotemporal) measurement and other observation data. ISO 19156 defines an application schema as a reference schema for data required by applications. O&M can be used as a generic means to deal with measurements and other observations in a standardized way. The O&M standard is developing over time and the newest version 3 that is under revision allows to add more specifics about measurements taken[110].

The specifications for an observation or measurement are based on the following main classes:

- the feature of interest (FoI on the instance level)
- the observable property of the FoI
- The observation itself
- The process/method used
- The resulting value (including the Unit of Measure)

In case of using a sample, some additional classes are modelled.

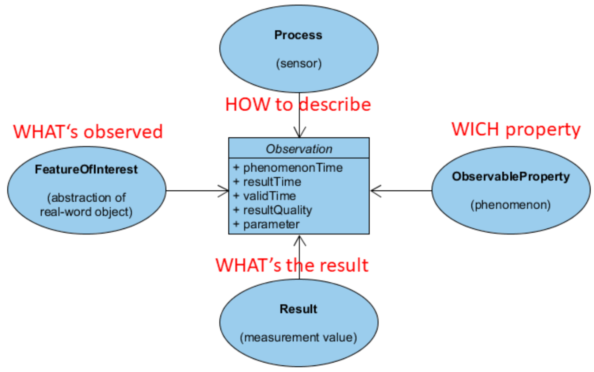
[110] https://www.ogc.org/projects/groups/omswg

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SENSORS AND STRUCTURED EXCHANGE DATA

OGC SensorThings API provides an open, geospatial-enabled and unified way to interconnect the Internet of Things (IoT) devices, data, and applications over the Web. The data exchange is based on the concepts of the O&M specification

SensorThings API Entities	O&M 2.0 Concepts
Thing (and Locations, HistoricalLocations)	-
Datastream	-
Sensor	Procedure
Observation	Observation
ObservedProperty	Observed Property
FeatureOfInterest	Feature-Of-Interest



SensorThings API Entity types	Relation	Description
Location	Many optional to many optional	The Location entity locates the Thing. Multiple Things MAY be located at the same Location. A Thing MAY not have a Location. A Thing SHOULD have only one Location. However, in some complex use cases, a Thing MAY have more than one Location representations. In such case, the Thing MAY have more than one Locations. These Locations SHALL have different encodingTypes and the encodingTypes SHOULD be in different spaces (e.g., one encodingType in Geometrical space and one encodingType in Topological space).
HistoricalLocation	One mandatory to many optional	A Thing has zero-to-many HistoricalLocations. A HistoricalLocation has one-and-only-one Thing.
Datastream	One mandatory to many optional	A Thing MAY have zero-to-many Datastreams.

Figure 3.6 alignment of concepts between O&M and SensorThings

SensorML is a self-contained and highly flexible data exchange format. This makes life easy for data producers but is demanding on consumers. SensorML provides extensive support for serialization of numeric data arrays and is particularly optimized for data that includes multiple parallel streams that must be processed together. For example, the data collected by cameras on airborne vehicles must be geo-referenced based on the instantaneous position of the platform and orientation of the camera. In contrast, **O&M** was designed to be more 'user-centric' with the target of the observation and the observed property as first-class objects. O&M works at a higher semantic level than SensorML, but only provides abstract classes for sensors, features of interest and observable properties, expecting the details to be provided by specific applications and domains. O&M also provided a model for sampling, since almost all scientific observations are made on a subset of, or proxy for, the ultimate feature of interest.

3.3.6 Ontologies for lists and vocabularies

To avoid confusion, first we give some background and definitions related to lists/vocabularies (vocabs).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

CONTROLLED VOCABULARY/CODE LIST/THESAURUS/TAXONOMY/ONTOLOGY

Controlled vocabularies provide a way to organize knowledge for subsequent retrieval (see Figure 3.10. They are used in subject indexing schemes, subject headings, thesauri, taxonomies and other knowledge organisation systems. Controlled vocabulary schemes mandate the use of predefined, authorised terms that have been preselected by the designers of the schemes, in contrast to natural language vocabularies, which have no such restriction.

A **Code list** is a form of controlled vocabulary containing a finite list of codes and meanings that represent the only allowed values for a particular data item.

In general context a **thesaurus** or synonym dictionary is a reference work for finding synonyms and sometimes antonyms of words. While in the dictionary you can see the word's definition and how it's used in speech (noun, verb, adjective etc.), when you want to know similar words you have to look in a thesaurus. In many cases also relations to other term are given like broader, narrower and related terms. And sometimes a thesaurus also includes words with opposite meaning, antonyms. In the context of information retrieval, a thesaurus (plural: "thesauri") is a form of controlled vocabulary that seeks to dictate semantic manifestations of metadata in the indexing of content objects. A thesaurus serves to minimise semantic ambiguity by ensuring uniformity and consistency in the storage and retrieval of the manifestations of content objects. ANSI/NISO Z39.19-2005 defines a content object as "any item that is to be described for inclusion in an information retrieval system, website, or other source of information". The thesaurus aids the assignment of preferred terms to convey semantic metadata associated with the content object. AGROVOC[111] is an example of a Thesaurus using SKOSMOS.

Taxonomy in 'general' is the practice and science of classification of things or concepts, including the principles that underlie such classification. Taxonomy in 'search engines' refers to classification methods that improve relevance in vertical search. Taxonomies of entities are tree structures whose nodes are labelled with entities likely to occur in a web search query. Searches use these trees to match keywords from a search query to keywords from answers (or snippets). Taxonomies, thesauri and concept hierarchies are crucial components for many applications of information retrieval, natural language processing and knowledge management. A number of tools using SKOS[112] standard are also available to streamline work with taxonomies. Taxonomy in 'soil', more specific for soil classification, deals with the systematic categorization of soils based on distinguishing characteristics as well as criteria that dictate choices. NCBI is an example of an on-line taxonomy database[113].

In computer science and information science, an **ontology** or domain ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many or all domains of discourse. More simply, an ontology is more than just a list, it is a way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject. The fundamental difference between an ontology and a controlled vocabulary is the level of abstraction and relationships among concepts. **A formal ontology uses an ontology representation language.** This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints (e.g., specifies what it means to be a well-formed statement, assertion, query, etc.) on how terms in the ontology can be used together. An example of an ontology is the W3C IoT ontology[114]

[Definitions source: Wikipedia]

Pieterse and Kourie (2014) describes that 'Knowledge organisation systems (KOS)' are mechanisms for organizing information, but many ambiguities exist when controlled vocabularies with reference to

---

[111] http://aims.fao.org/vest-registry/vocabularies/agrovoc ; http://skosmos.dev.finto.fi/agrovoc/en/page/c_3219

[112] SKOS, which stands for Simple Knowledge Organization System, is a W3C standard, based on other Semantic Web standards (RDF and OWL), that provides a way to represent controlled vocabularies, taxonomies and thesauri. Specifically, SKOS itself is an OWL ontology and it can be written out in any RDF syntax

[113] National Center for Biotechnology Information: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2599362

[114] https://semwebtec.wordpress.com/2010/11/23/contolled-vocabulary-vs-ontology/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

classification-related terms are used. Due to the need to organize information in all disciplines, knowledge organisation systems (KOSs) with varying attributes, content and structures have been created in different domains. Hodge (2000) describes a classification of KOSs based on characteristics such as structure and complexity, relationships between them, and historical functions. This classification is presented in Figure 3.9.

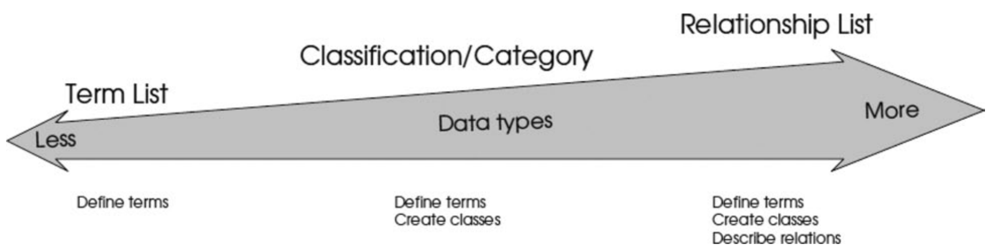


Figure 3.7 Increasing data types included in KOSs

Gilchrist (2003) clarifies the distinction between different KOSs. He proposes the progression on content types shown in Figure 3.10.

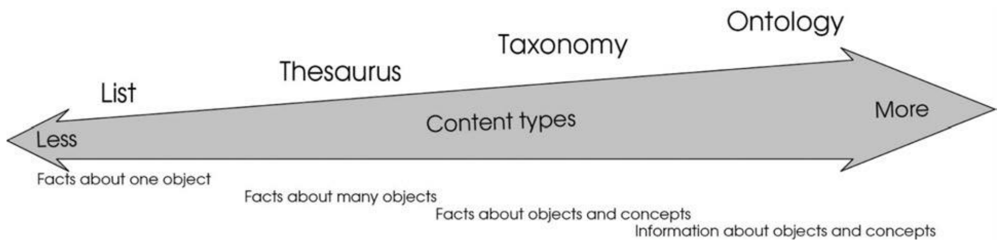


Figure 3.8 Increasing types of content included in KOSs

In the classification of KOSs of Pieterse and Kourie the inherent structure of classifications is considered. Classes of KOSs are characterized by the progressive addition of features that enhance the capabilities offered by these KOSs. The addition of these features contributes to their increased complexity. We call these classes of KOSs lists, taxonomies, lattices[115], thesauri and ontologies. This progression is summarized in Figure 3.11.

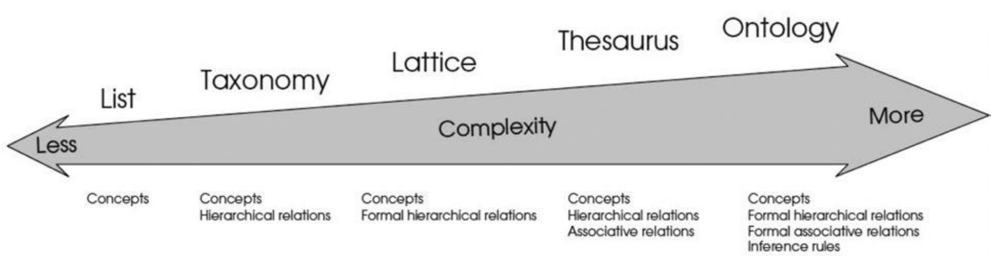


Figure 3.9 Increasing structural complexity of KOSs

[115] Lattices are originating from mathematics and not considered here

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

The increasingly complexity reflects from simple code lists at the left hand side to full domain ontologies at the right side.

In soil often classifications are used, which basically are also controlled vocabularies just like code lists. The "systematic" classification such as for example the soil classification referred to in the definition of taxonomies is one type, but also classification rules to define intervals or ranges in fact groups individual values into 'classes', or sometimes named as 'categories', according to 'established' intervals (for example for texture). So for the sake of clarity we would like to refer to these as 'interval' lists here. This is again a type of controlled vocabulary.

An important notion is to differentiate between code lists and so-called 'property' lists. The code lists provide values to be used for attributes, the properties of the classes that are describing the model (or attributes). A list of properties does not contain values but is meant to list what properties exist and can be added as attributes to objects in the model. In general, these are 'observable' properties and can be best described following the O&M specifications. Chemical and physical properties of a soil Layer (horizon) can illustrate this. For physical and chemical properties, a full list of all possible items that are relevant for the soil domain exists/is made. For a specific (use) case we only want to use a 'subset' of the full list of properties we really need (or to exchange). When we want to calculate the Soil Organic Carbon (SOC) stocks, we need soil organic carbon concentration, soil bulk density, fine earth fraction (given by 1-coarse fragment content) and the thickness of the soil layer to calculate the SOC stock. For this case, the property list contains of the chemical properties we need, so the items on the chemical property list: Soil Organic Carbon, for physical properties the bulk density, the coarse fragment content; (and perhaps the clay, silt and sand distribution if bulk density needs to be calculated with a pedotransfer function). These are the subsets to be used from the full lists and it should be decided if these are separate code lists or integrated into the full list, which is more complex and probably more difficult to maintain. Property lists are just to be seen as controlled vocabularies and act as a placeholder for what properties could or need to be included in the model for that case (or included when exchanging data). Those lists do not reflect relations that exist in the model. For instance in O&M[116], where a measurement is modeled, a certain method that is used for that measurement is a separate class in the model and has its own code lists. As an example, to measure the calcium carbonate content, possible methods are: the titration, inductively coupled plasma (ICP), X-ray diffraction (XRD) *TOPAS*, thermogravimetric analysis (TGA), ASTM, and washing methods. The $CaCO_3$ is as a chemical property in the property codelist and the methods used to measure $CaCO_3$ are in another codelist possibly named 'CalciumCarbonateMethods' and those two are separate controlled vocabs, related but not explicitly linked.

[116] ISO 19156: Observations & Measurements (O&M)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

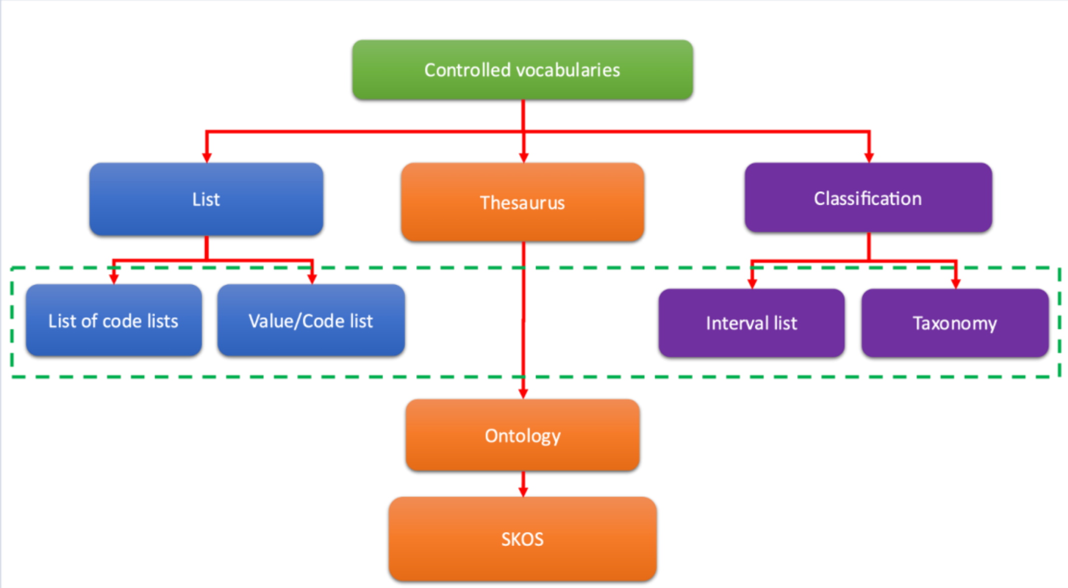


Figure 3.10 -Summary overview of types of lists/vocabularies

As a summary, basically all types of lists are to be referred to as controlled vocabularies (overview in Figure 3.10). However, different structures can be needed to provide the proper information. The challenge is to add not only the data itself, the listed items, but also all the metadata of those items to find, access and (re-)use them in a defined consistent (unambiguous) way to allow not only human but also machine-to-machine use in an automated manner.

3.4 Controlled vocabularies and ontologies on Soil

3.4.1 Soil controlled vocabularies

Soil properties are included in several controlled vocabularies, ranging from simple code lists up to more complex thesauri including relationships between terms. Also referred to as ontologies, but not meant as being a 'domain ontology'. They are accessible on AgroPortal[117], Bioportal[118], Ontoportal[119], VEST Registry[120] , or the Ontology Lookup Service[121]. Among those, some resources are very soil specific or include a range of soil related properties, while others are more general, but provide concepts commonly used to describe soils (e.g. chemical composition). They also differ in terms of richness and comprehensiveness and sometimes new resources are developed integrating and enlarging previous ones. An example of a soil-related ontology could be ECSO (Ecosystem Ontology). According to a survey carried out by the Soil Ontology and Informatics Cluster[122], its comparison with other ontologies

117 http://agroportal.lirmm.fr/
118 https://bioportal.bioontology.org/
119 https://ontoportal.org/
120 https://vest.agrisemantics.org/vocabularies?combine=soil&field_original_source_target_id=All&tid=All&term_node_tid_depth=All&field_vocabulary_type_target_id=All&field_type_of_entity_target_id=All&page=1
121 https://www.ebi.ac.uk/ols/index
122 https://esipfed.github.io/soil_data_model_survey/index.html

revealed it performs the best for soil topics. A list of knowledge representation tools, is presented in Table 3.1 below.

*Table 3.1 Soil controlled vocabularies.. Type: **O**=ontology; **T**= thesaurus; Topic: **S**= soil specific; **R**=including soil related terms. Resources URI in footnotes. Metrics information have been retrieved through origin portal, Protégé or WebVOWL.*

Name - URI	Type	Topic	Description	References	Format	Metrics
WRB 2014[123]. Update 2015	O	S	World Reference Base for Soil Resources v. 2014 (update 2015) – qualifiers	L'Abate et al., 2016; L'Abate et al., 2017; Caliper-FAO, 2020	RDF/ XML	4,854 Axioms; 4 Classes; 300 Individuals. 276 results for 'soil'.
agINFRA Soil Vocabulary [124]	O	S	Ontology implementation of the INSPIRE Soil model	L'Abate et al., 2015	RDF OWL	146 Axioms; 18 Classes; 10 Object Properties; 1 Data Properties; 21 Annotation Properties. 12 results for 'soil'.
Soil-Property-Process[125]	O	S	It mainly describes soil properties and processes and their interactions	Du et al., 2016	OWL	3,916 Axioms; 592 Classes; 16 Object Properties (C-BY 4.0). 6 results for 'soil'.
NCBITAXON [126]	O	R	NCBI organismal classification, an ontology representation of the NCBI organismal taxonomy	Midford et al, 2013	OWL	Not available (failed to load ontology). 2905 results for 'soil'.
ENVO[127]	O	R	Ontology of environmental features and habitats. Using ENVO to describe things like ecosystems, entire planets and other astronomical bodies, their parts, or environmental processes increases the interoperability of environmental descriptions, helping (meta)data records achieve demonstrable FAIRness	Buttigieg et al., 2016	OWL	45,864 Axioms; 6,199 Classes; 134 Object properties; 1 Data properties; 44 Individuals; 101 Annotation Properties. 369 results for 'soil'.

123 https://stats-class.fao.uniroma2.it/caliper/sites/default/files/classifications/WRB2014_update2015.rdf
124 http://data.agroportal.lirmm.fr/ontologies/SOIL/submissions/1/download?apikey=1de0a270-29c5-4dda-b043-7c3580628cd5
125 http://imash.leeds.ac.uk/ontologies/atu/SoilPhysics.owl#
126 http://purl.obolibrary.org/obo/ncbitaxon.owl
127 http://purl.obolibrary.org/obo/envo.owl

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Name - URI	Type	Topic	Description	References	Format	Metrics
DEMETER[128]	O	R	The DEMETER Agriculture Information Model (AIM) is the common vocabulary in the DEMETER project providing the basis for semantic interoperability across smart farming solutions	Roussaki et al., 2019	OWL CSV RDF/ XML	3021 Axioms; 174 Classes; 154 Object properties; 124 Data properties; 127 Individuals; 27 Annotation properties. 230 results for 'soil'.
ChEBI[129]	O	R	Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds	Degtyarenko et al., 2008	OWL OBO CSV RDF/ XML Diff	155,627 Classes; 0 Individuals; 10 Properties. 40 results for 'soil'.
ECSO[130]	O	R	The Ecosystem Ontology, DataONE ontology of Carbon Flux measurements for MsTMIP and LTER Use Cases	Mecum, 2019	OWL CSV RDF/ XML Diff	19,667 Axioms; 2,094 Classes; 120 Object properties; 2 Data properties; 18 Individuals; 49 Annotation Properties. 25 results for 'soil'.
agroRDF[131]	O	R	agroRDF is a Semantic Overlay to agroXML. AgroXML is an XML dialect for representing data on work processes on the farm, including accompanying operating supplies like fertilizers, pesticides, crops and the like.	Martini et al., 2013	OWL CSV RDF/ XML	189 Classes; 24 Individuals; 107 Properties. 8 results for 'soil'.
AGRO[132]	O	R	AgrO, the Agronomy Ontology, describes agronomic practices, techniques, and variables used in agronomic experiments	Arnaud et al., 2020	OWL	20,647 Axioms; 2,268 Classes; 185 Object Properties; 6 Datatype Properties; 1,239 Individuals; 6 results for 'soil'.

[128] https://w3id.org/demeter/agri (last update 28/10/2020)
[129] http://purl.obolibrary.org/obo/chebi.owl
[130] http://purl.dataone.org/odo/ECSO8.owl
[131] http://data.igreen-services.com/agrordf
[132] http://data.agroportal.lirmm.fr/ontologies/AGRO/download?apikey=1de0a270-29c5-4dda-b043-7c3580628cd5&download_format=rdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Name - URI	Type	Topic	Description	References	Format	Metrics
QUDT[133]	O	R	QUDT - Quantities, Units, Dimensions and Data Types Ontologies	Rijgersberg et al., 2013	OWL	6,495 Axioms; 345 Classes; 191 Object properties; 129 Data properties; 169 Annotation properties. 0 results for 'soil'.
OM[134]	O	R	The Ontology of units of Measure (OM) 2.0 models concepts and relations important to scientific research. It has a strong focus on units, quantities, measurements, and dimensions.	Rijgersberg et al., 2021 (2.0.29)	OWL CSV RDF/ XML	23,568 Axioms; 812 Classes; 23 Object properties; 11 Data properties; 2,212 Individuals; 39 Annotation Properties. 0 results for 'soil'.
EARTh[135]	T	R	An Environmental Application Reference Thesaurus in the Linked Open Data Cloud	Albertoni et al., 2014	RDF	251,386 Axioms; 1 Classes; 0 Object properties; 0 Datatype properties; 14,351 Individuals; 15 Annotation Properties. 855 results for 'soil'.
AnaEE THES[136]	T	R	The AnaEE thesaurus aims to provide a controlled vocabulary for the semantic description of continental ecosystems and their biodiversity	Clastre et al., 2018	SKOS/ RDF	2 Classes; 0 Object properties; 0 Datatype properties; 3247 Individuals. 748 results for 'soil'.
NALT[137]	T	R	The Thesaurus is an online vocabulary of agricultural terms in English and Spanish	Developed by U.S. Department of Agriculture[138]	SKOS/ RDF	817,729 Axioms; 2 Classes; 76,504 Individuals; 15 Annotation Properties. 629 results for 'soil'.
GEMET[139]	T	R	The GEneral Multilingual Environmental Thesaurus, has been developed as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data	Developed by Eionet[140]	RDF/ XML	325,116 Axioms; 8 Classes; 0 Object Properties; 0 Data Properties; 5,695 Individuals; 34 Annotation Properties. 392 results for 'soil'.

133 http://www.qudt.org/2.1/schema/qudt
134 http://www.ontology-of-units-of-measure.org/resource/om-2/
135 http://purl.oclc.org/net/DumpEarthRDF
136 https://opendata.inra.fr/anaeeThes/page/ (last update 19/03/2020)
137 https://agclass.nal.usda.gov/sites/default/files/2021/NAL_Thesaurus_2021_SKOS.zip (last update 28/01/2020)
138 https://agclass.nal.usda.gov/
139 http://data.agroportal.lirmm.fr/ontologies/GEMET/download?apikey=1de0a270-29c5-4dda-b043-7c3580628cd5&download_format=rdf
140 https://www.eionet.europa.eu/

Name - URI	Type	Topic	Description	References	Format	Metrics
			Sources (ETC/CDS) and the European Environment Agency (EEA), Copenhagen			
GACS[141]	T	R	Integrates three thesauri: the AGROVOC Concept Scheme, the CAB Thesaurus (CABT), and NAL Thesaurus (NALT)	Baker et al., 2019	RDF/ XML TURTLE JSON-LD	3,894 Axioms; 6 Classes; 222 Individuals; 14 Annotation Properties. 149 results for 'soil'.
AGROVOC[142]	T	R	It is the largest Linked Open Data set about agriculture available for public use and facilitates access and visibility of data across domains and languages	Caracciolo et al., 2013	RDF/ XML TURTLE JSON-LD	748 Axioms; 3 Classes; 43 Individuals; 38 Annotation Properties. 139 results for 'soil'.

Table 3.2 Prototype ontology GLOSIS

Name - URI	Type	Topic	Description	References	Format	Metrics
GloSIS Ontology[143]	O	S	A prototype Ontology for GloSIS. This ontology is meant solely for development purposes. It does not yet intend to convey an actual domain model for soil information	De Sousa, 2020	TTL OWL	37 Axioms; 10 Classes; 9 Object Properties; 13 Data Properties; 0 results for 'soil'.

3.4.2 Discussion on listed controlled vocabularies

After this first review of the concepts, development and the status of possibilities for structuring soil data into ontologies, there has been a lot of development on semantics in the recent past, also within the soil discipline. An important step forward seems to be achieved by the H2020 SIEUSOIL project, developing the GloSIS Ontology[144] which was exposed in several, under deployment versions. The ontology builds on other relevant standards, or de facto standard ontologies in the soil domain. Among the other listed ontologies that could be reviewed for further ontology development are the WRB2014, update 2015; Soil-Property-Process, to include a wide variety of soil processes; ChEBI, which lists several chemicals also analysed within the Soil domain; QUDT, to help managing Unit of measurements and Quantities, Units, Dimensions and Data Types. Other cited ontologies might also contribute to

141 http://browser.agrisemantics.org/rest/v1/gacs/data?uri=http%3A%2F%2Fid.agrisemantics.org%2Fgacs%2FG_JJ&format=application/rdf%2Bxml (last update 13/12/2017)
142 http://aims.fao.org/aos/agrovoc/ (last update 02/02/2021) ;
143 https://git.wur.nl/duque004/glosis-owl/-/blob/master/GloSIS.owl
144 https://github.com/rapw3k/glosis

improve a soil domain specific ontology for some specific portions of semantics based on the INSPIRE Soil conceptual model.

3.4.3 A domain ontology for soil: INSPIRE Soil

For the purpose of creating an ontology we define soil, as done for example in INSPIRE, as 'the upper part of the earth's crust, formed by mineral particles, organic matter, water, air and living organisms. It is the interface between rock, air and water which hosts most of the biosphere.' With this definition the scope covers:

- Soil inventories, providing one-off assessments of soil conditions and/or soil properties at certain locations and at a specific point in time, and allow soil monitoring, providing a series of assessments showing how soil conditions and/or properties change over time.
- Soil mapping, providing a spatial presentation of the properties linked to the soils, including soil types; typically, soil maps are derived with the help of data available in soil inventories. Also other soil related information derived from soil properties, possibly in combination with non-soil data are within the scope.

The ontology contains a core set of spatial object types and their attributes that are considered to be essential for the infrastructure along which data on soil can be exchanged.

On the basis of the previous considerations, the soil theme includes the following phenomena or features of interest (either in the physical world or conceptualized world):

- soil profiles
- soil sites, soil plots
- soil bodies (delineated areas on the earth's surface determined on the basis of certain soil characteristics)
- soil characteristics (parameters) that change over time

This is depicted in Figure 3.11 and Figure 3.12:

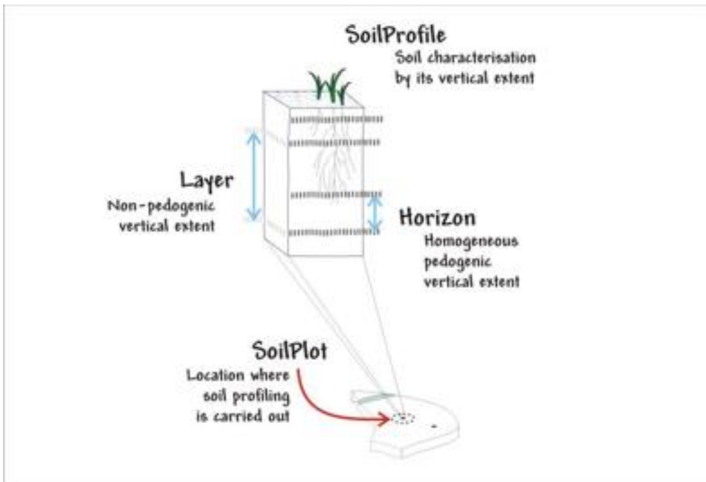


Figure 3.11 Soil profile, layer and Horizon located in a soil plot (INSPIRE data specifications)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Used as core elements to be mapped spatially that can be depicted as follows:

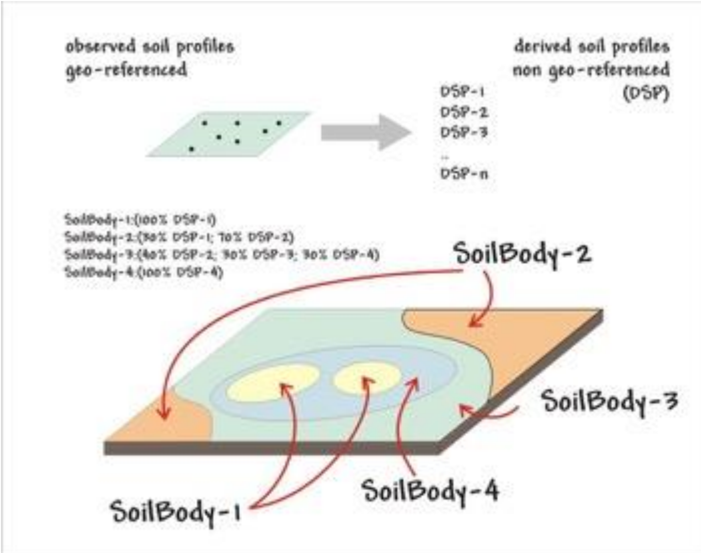


Figure 3.12 Core elements mapped spatially as soil delineated areas into soil bodies (INSPIRE data specifications)

As such, all entities have been identified, but the next step is to describe them with their relationships. Describe means adding definitions and associated properties to the entities with again their definitions and allowed values. All of this is based on knowledge of domain experts who provide each their own ontology fitting within the domain ontology. Domain ontology describes the whole encompassed by the envisioned universe of discourse. In the next section the main features of the soil ontology are described.

3.4.4 The INSPIRE Soil Data specification

The INSPIRE domain model[145] as described above is specified in the INSPIRE Consolidated UML Model[146], with its underlying INSPIRE XML schema according to the ISO standards (ISO/TS 19103:2005:Geographic information - Conceptual schema language and ISO 19109 the Generic Conceptual Model. In Figure 3.13 we give a simplified representation of the Consolidated UML Model.

To explain the INSPIRE UML model for the theme soil, we use the descriptions given in the INSPIRE data specifications and we focus the soil application schema, where the following features of interest are included derived from phenomena that already been stated in the previous section:

- SoilProfile (including ObservedSoilProfile and DerivedSoilProfile)
- ProfileElement (including SoilLayer and SoilHorizon)
- SoilBody
- SoilDerivedObject
- SoilThemeCoverage and SoilThemeDescriptiveCoverage
- SoilSite

[145] https://inspire.ec.europa.eu/Themes/127/2892

[146] https://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/index.htm?goto=2:3:17:1:8708 , version 1.0 - Keijerst, 2012

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

- SoilPlot

In the simplified soil theme UML model of Figure 3.13 the SoilThemeCoverage and SoilThemeDescriptiveCoverage are not reported, instead they are present in the Consolidated UML model

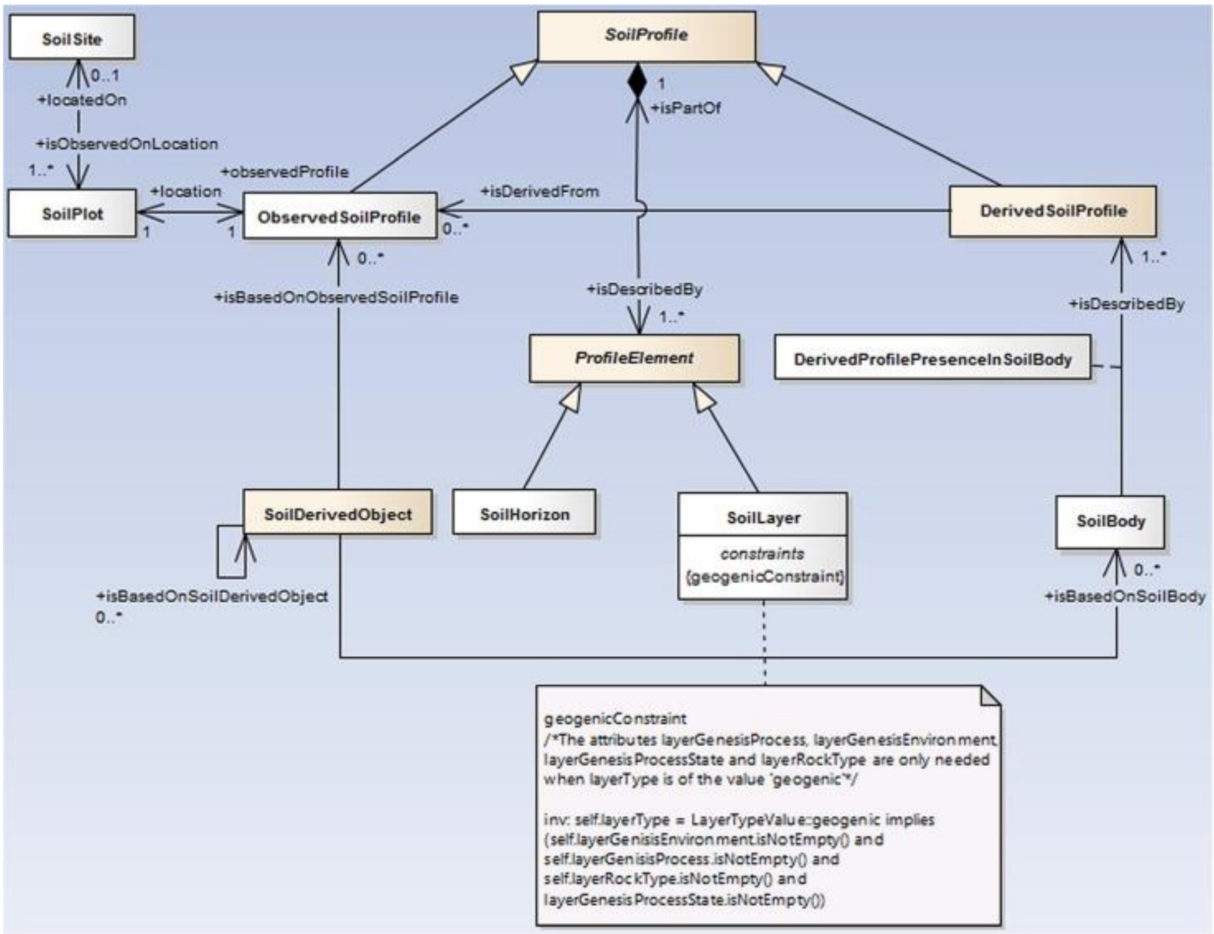


Figure 3.13 simplified UML INSPIRE soil application schema depicting the features of interest and it relationships (classes)

In the data specifications the features of interest are described as follows.

Soil Profile

For SoilProfile a distinction is made in an ObservedSoilProfile and a DerivedSoilProfile. An ObservedSoilProfile represents a geo-referenced soil profile, described in the field, possibly sampled and analysed in the laboratory. A DerivedSoilProfile is a non-point-located SoilProfile with property values that are derived (e.g. averaged) from the values of the corresponding properties of one or more ObservedSoilProfiles. Even if such a connection to an ObservedSoilProfile exists, it is not mandatory to provide it (nor its data) together with the DerivedSoilProfile.

Any soil profile can be characterized as a whole by a number of properties, of which the following are included in the model: its soil type according to the WRB soil classification scheme (WRBSoilName) and/or any other soil classification scheme (otherSoilName) with the limitation to one per dataset, and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

zero or more other parameters, which are expressed through soilProfileObservation associations with OM_Observation objects. Through the observedProperty role of its Phenomenon association, the OM_Observation object designates the parameter, through the attributes label, basePhenomenon (value selected from the codelist SoilProfileParameterNameValue) and UoM (unit of measure). Through the result role of the Range association, a value can be given to the parameter; this value should be of the type Number, RangeType (a range of values) or CharacterString (e.g. 'good' or 'very high'). Note that the SoilProfileParameterNameValue codelist can be extended by the data provider when needed

Taxonomic Classes, and Soil Typological Units. The concepts behind the DerivedSoilProfile

Taxonomic classes are concepts. Each taxonomic class has a set of soil characteristics with precisely defined limits. The classes are used as a basis for comparison to classify soils systematically. Taxonomic classes summarize an immense amount of research and experience related to the significance of soil properties and combinations of properties. They provide predefined sets of soil properties that have been tested for genetic relationships and interpretative value.

Taxonomic classes are at the base of the soil correlation process (chapter 5.3) which leads to the production of soil types, or so-called **Soil Typological Units (STU)**. In the INSPIRE conceptual model the STU concept has been better defined by the concept of the DerivedSoilProfile, as described above.

ProfileElement (SoilLayer and SoilHorizon)

Any soil profile, whether observed or derived, can be described by horizons and layers. Each horizon and layer can have a number of properties. In the model, layers and horizons are represented by the classes SoiLayer and SoilHorizon which are both subtypes of the abstract class ProfileElement. The abstract SoilProfile can consist of one or more ProfileElements. A horizon or layer is at least characterized by an upper depth and a lower depth, indicating the top and the bottom depth of the horizon or layer from the surface; the attribute in the abstract ProfileElement class that indicates the depths of a horizon or layer is profileElementDepthRange.The properties of horizons and layers are modelled through the profileElementObservation associations with OM_Observation objects, in the same way as soil profile parameters are modelled, the only difference being that the parameter is selected from the codelist ProfileElementParameterNameValue. Note that this codelist can be extended by the data provider when needed. A horizon is further specified by a horizon name according to the FAO horizon notation scheme from 2006 (FAOHorizonNotation) and/or any other horizon notation schemes (otherHorizonNotation), with the limitation to one per a dataset. A horizon corresponds to a horizontal subdivision of the soil based on pedogenic processes. A SoilLayer corresponds to a horizontal subdivision of the soil based on other criteria than pedogenic processes. The way of defining a layer is specified by a layer type name that indicates the kind of layer considered: topsoil, subsoil, depthInterval or geogenic; this is modelled through the layerType attribute in the SoilLayer class. Topsoil and subsoil are complementary concepts used to address pedogenic process domains of the soil irrespective of a horizon description. Depth intervals are often used for chemical characterisation of the soil state and relate often to sampling depths. If the SoilLayer is of the type geogenic, it is described in terms of its non-pedogenic origin and can additionally be described by the following attributes: layerRockType (which gives petrographic or lithologic information on the rock type the layer is made of) and three attributes with reference to layer genesis (layerGenesisProcess, layerGenesisEnvironment, layerGenesisProcessState). Except for layerGenesis-ProcessState, the involved codelists originate from the INSPIRE Data Specification on Geology. Note that the values in the codelists SoilProfileParameterNameValue and ProfileElementParameterNameValue in this version of the model correspond to requirements concerning soil property data in some European legislation. Note that, since the parameters for soilProfile, ProfileElement and soilDerivedObject are linked (through the O&M framework) to OM_Observation, which in turn is associated to INSPIRE_OM_Process, it is possible to provide additional information on the process that led to observation values. For example, if soil pH is measured in a salt solution, the kind and concentration of the salt solution as well as the solution to soil proportion and the type of device used can be stated.

SoilBody

To delineate geographically areas with a soil cover that can be characterized by a set of derived soil profiles, the model introduces the construct of the SoilBody class. It represents an association (or other types of spatial linkages of various soil types) of derived soil profiles that represent the soils found together in the area of the SoilBody. The area is specified by the geometry attribute of the SoilBody. The presence of one or more kinds of soils in the SoilBody is modelled with the association class DerivedProfilePresenceInSoilBody, which allows to indicate which derived soil profiles are used to describe the soils of the SoilBody, and to which extent (expressed as a couple of area share percentages). The couple of percentages offer the flexibility to give a range of percentages to express uncertainty on the presence of any soil type. If only one percentage value is to be used, lower and upper boundaries of the couple of percentages should have identical values. Because of this flexibility with ranges of percentages, it is allowed that the sum of all percentage upper boundaries for the derived soil profiles in one soil body is greater than 100%. However, there is the constraint that the sum of all percentage lower boundaries for the derived soil profiles in one soil body is lower than or equal to 100%. As an example, a SoilBody could consist of one dominant soil (as described by a derived soil profile) and of other soils (described by other derived soil profiles) having characteristics different from the dominant one. A derived soil profile can be used to characterize more than one SoilBody. The soilBodyLabel attribute of the SoilBody allows a description of the SoilBody, which may be useful for building legends. The soilBodyLabel contributes to the explanation of a mapping unit of a map, whereas in the metadata linked to the dataset to which the object belongs, a reference should be given to documentation that further explains the labelling of the soil bodies.

SoilDerivedObject

In the context of the model, a SoilDerivedObject is defined as a spatial object (e.g. a point, line, polygon) representing a soil-related property (using the association soilDerivedObjectObservation with an OM_Observation object) which value can be (but does not have to be) derived from a) values of soil properties of related observed soil profiles and/or related soil bodies, and/or b) any other data or information intern or extern to the model (for example: instances of other SoilDerivedObjects (intern); landcover/climate data (extern)). A collection of such SoilDerivedObjects constitutes a soil thematic map and is to be regarded as a dataset. The metadata linked to such a dataset provides the details on how the values for the attributes of the SoilDerivedObjects have been calculated. In the INSPIRE model this is an 'open' Class, meaning no real properties, beside an Id and geometry are attached to have meaningful instances of that class (observable) properties have be added (possibly defined in extentions).

Soil Theme Coverage, Soil Theme Descriptive Coverage

The specific purpose of the SoilThemeCoverage class is to provide a structure for the interchange of soil thematic maps as continuous phenomena. The geometry is represented as a coverage which is defined as a ― feature that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain‖. For soil this commonly is a rectified grid. Since it may be useful to associate to a coverage other coverages of which the cell values are supplementary information to the corresponding grid cells of the coverage itself, the SoilThemeCoverage class has an association to another coverage, the SoilThemeDescriptiveCoverage. An example would be the provision of ― purity maps in relation to for instance soil texture maps.

SoilSite, SoilPlot

A SoilSite is considered as a wider geographical area, i.e. the larger piece of land where soil investigation takes place in one or more spots, called soil plots. A site represents often just the geographically not strictly defined environment of the plots; thus, the geometry attribute of the soil site can be a surface or a point location. In soil survey and general soil monitoring, all soil information gathered on one site is handled as if it would have been collected at the very same location, which is impossible in the strict sense in the real world whenever soil investigation is destructive. The purpose of this investigation can be general (e.g. taking soil profiles for a general soil characterization) or specific (e.g. sampling to investigate potentially contaminated land). The SoilPlot object is included in the model to provide the type and location of the associated observed soil profile. A SoilPlot within a SoilSite is of a certain type (borehole, sample, trial pit) and located by a geographical point and/or the name of a location. To a SoilPlot , one ObservedSoilProfile must be associated. A SoilSite is represented in the model with the SoilSite class. Its soilInvestigationPurpose attribute indicates the purpose of investigation: general (generalSoilSurvey) or specific (specificSoilSurvey). This kind of information can be crucial for data evaluation to identify bias in the selection of sites. The possible properties of a soil site are modelled through the soilSiteObservation associations with OM_Observation objects, in the same way as soil profile parameters are modelled, the only difference being that the parameter is selected from the codelist SoilSiteParameterNameValue. Note that this codelist can be extended by the data provider when needed. A soil plot is represented in the model with the SoilPlot class. A SoilSite comprises one or more SoilPlot-s. A soil plot is of a certain type (soilPlotType), and its location is indicated by the attribute soilPlotLocation which can take the form of either a specific X,Y-location or a description of the location using text or an identifier.

For more information and details, see the INSPIRE data specifications (Footnote 146).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

3.4.5 The INSPIRE Registry

The INSPIRE infrastructure involves several features, which require clear descriptions and the possibility to be referenced through unique identifiers. Examples for such features include INSPIRE themes, code lists, application schemas or discovery services. Registers provide a means to assign identifiers to items and their labels, definitions and descriptions (in different languages). The INSPIRE registry provides a central access point to a few centrally managed INSPIRE registers. The content of these registers is based on the INSPIRE Directive, Implementing Rules and Technical Guidelines. It is powered by: Re3gistry 1.3 software, which was being developed by the Are3NA project as part of the Interoperability Solutions for European Public Administrations (ISA) Programme.

3.4.6 Codelist in INSPIRE[147]

An overview of the current (March 2021) codelists in the INSPIRE codelist registry is provided here:

FAO Horizon Subordinate (FAOHorizonSubordinateValue) - A code list of designations of subordinate distinctions and features within the master horizons and layers which are based on profile characteristics observable in the field and are applied during the description of the soil at the site. Not extensible. Values are not yet exposed into the INSPIRE registry;

FAOPrimeValue (FAOPrimeValue) - A prime and double prime may be used to connotate the master horizon symbol of the lower of two (prime) or three (double prime) horizons having identical Arabic-numeral prefixes and letter combinations. Not extensible. Values are not yet exposed into the INSPIRE registry.

LayerGenesisProcessStateValue (LayerGenesisProcessStateValue) - An indication whether the process specified in layerGenesisProcess is ongoing or has ceased. Not extensible. Coded list values (2 entries);

LayerTypeValue; (LayerTypeValue) - A classification of a layer according to the concept that fits the purpose. Not extensible. Coded list values (4 entries);

OtherHorizonNotationTypeValue; (OtherHorizonNotationTypeValue) - A classification of a soil horizon according to a specific classification system. Empty code list;

OtherSoilNameTypeValue; (OtherSoilNameTypeValue); - An identification of the soil profile according to a specific classification schema. Empty code list;

Profile Element Parameter Name (ProfileElementParameterNameValue) - Properties that can be observed to characterize the profile element. Extensible with narrower values. Coded list values (12 entries);

SoilDerivedObjectParameterNameValue; (SoilDerivedObjectParameterNameValue); Soil-related properties that can be derived from soil and other data. Extensible with narrower values. Coded list values (17 entries);

SoilInvestigationPurposeValue; (SoilInvestigationPurposeValue); - A code list of possible values indicating the reasons for conducting a survey. Not extensible. Coded list values (2

[147] https://inspire.ec.europa.eu/codelist/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

entries);SoilPlotTypeValue; (SoilPlotTypeValue); A code list of terms specifying on what kind of plot the observation of the soil is made. Not extensible. Coded list values (3 entries);

SoilProfileParameterNameValue; (SoilProfileParameterNameValue); Properties that can be observed to characterize the soil profile. Extensible with narrower values. Coded list values (7 entries);

SoilSiteParameterNameValue; (SoilSiteParameterNameValue); - Properties that can be observed to characterize the soil site. Extensible with narrower values. Coded list values (136 entries);WRBQualifierPlaceValue; (WRBQualifierPlaceValue); A code list of values indicating the placement of the Qualifier with regard to the WRB reference soil group (RSG). The placement can be in front of the RSG i.e. 'prefix' or it can be behind the RSG i.e. 'suffix'. Not extensible. Values are not yet exposed into the INSPIRE registry;

WRBQualifierValue; (WRBQualifierValue); A code list of possible qualifiers of the World Reference Base for Soil Resources. Not extensible. Values are not yet exposed into the INSPIRE registry;

WRBReferenceSoilGroupValue; (WRBReferenceSoilGroupValue); list of possible reference soil groups (i.e. first level of classification of the World Reference Base for Soil Resources). The allowed values for this code list comprise only the values specified in World reference base for soil resources 2006, first update 2007. Not extensible. Coded list values (32 entries);

WRB specifiers (WRBSpecifierValue) - list of possible specifiers. SOURCE World reference base for soil resources 2006, first update 2007, World Soil Resources Reports No. 103, Food and Agriculture Organisation of the United Nations, Rome, 2007. Specifiers are name elements in WRB restricting the meaning of qualifiers. Not extensible. Values are not yet exposed into the INSPIRE registry.

3.5 Data harmonisation (method and lab)

As mentioned in paragraph 3.1, harmonisation provides mechanisms and advantages (in time, cost, understanding and communication) for the collation, analysis, and exchange of information as well as consistent and comparable global soil data. It concerns all the relevant issues of soil data, from survey to laboratory analysis, and interpretation to data exchange Figure 3.14.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

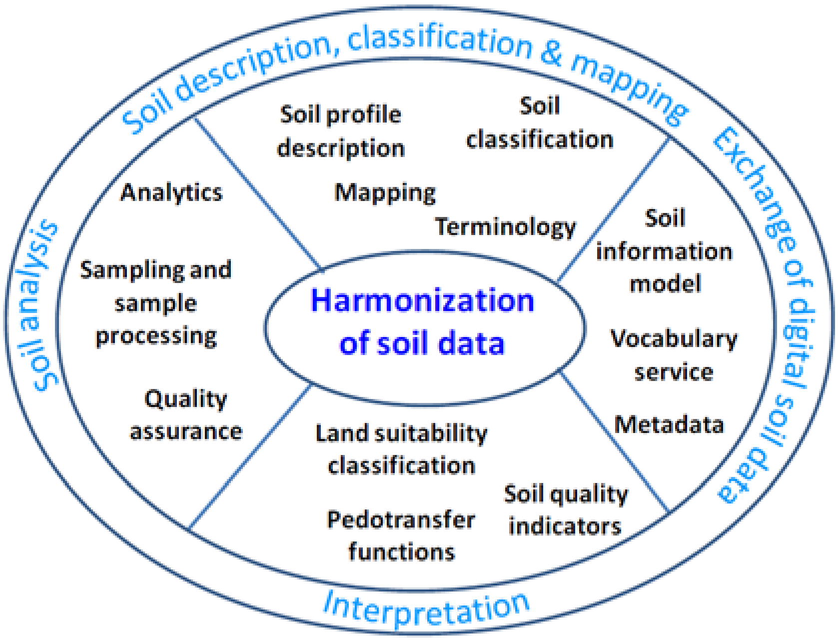


Figure 3.14 Data harmonisation: Key areas and topics (source: Baritz et al., 2014)

Soil profile description and classification

The basic soil information is often derived from soil profile descriptions. And standardised or harmonised soil profile description is the first topic when dealing with harmonisation of the soil information workflow or domain. Some general guidelines exist (see paragraph 3.1). The FAO guide has served as reference in many harmonisation procedures, even if due to recent improvements of relevant reference material this guideline requires updating. The current ongoing activities for revising and improving soil (profile) description is done in the IUSS working group 'Universal Soil Classification', and has synergies with GSP Pillar 5 activities. They aim to address the diverse areas for field profile descriptions, horizon nomenclature, designations and definitions.

Much of the information stored in legacy databases refers to national soil classification systems and soil profile description criteria. There are many examples of correlating the diverse national and international classification systems. These often differ conceptually as reviewed by Krasilnikov et al. (2009), creating an additional source of uncertainty.

As indicated, the World Reference Base for Soil Resources (WRB) and Soil Taxonomy are widely used for soil classification, scientific publication, and soil mapping on various scales, and are commonly used for correlations (Brevik et al., 2016). At national level, these are usually used in addition to the national classification nomenclature. The WRB classification has been chosen as a standard in several harmonisation projects and initiatives, and it is also retained by the INSPIRE soil thematic group, together with the FAO horizon notation scheme, with the technical and scientific support of ESDAC. However, the long-term project of defining a Universal Soil Classification System, uniting and improving the two, is on its way, with the support of the IUSS and the GSP. The underpinning concept is based on a data centroid approach (Hempel et al., 2013), used as a tool of correlation of soil units from different soil classification systems (Láng et al., 1013).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

The procedures for data model and exchange protocols are described in paragraphs 3.2and 3.3, while the soil sampling and the mapping harmonisation are addressed in chapters 4 and 5, respectively.

Lab standards - general

The ISO standards for soil analysis covers many soil chemical and physical analyses, including sampling and sample pre-treatments. However, many laboratories do not follow the ISO standards or interpret them differently due to several reasons such as cultural, limited infrastructure facilities, having their own well established reference methods (e.g. USDA NCSS)and costs. Moreover, many legacy data exist that were sampled and analysed with different methods in different periods.

In this context, the Plan of Action for Pillar 5 of the GSP[148] outlines possible strategies for overcoming these issues, when ISO standard cannot be applied. The strategies are listed as: developing conversions towards the standard methods, building on comparative analysis (e.g., analysis of archived samples), participating in inter-laboratory comparisons to develop conversions, and archiving samples for post re-analysis. These are elaborated further and executed in the workplans and objectives of GLOSOLAN (paragraph 1.2.9 and 3.1). GLOSOLAN adds to these the development of harmonised Standard Operating Procedures for often used methods, capacity improvement for laboratories on quality assurance and control and promoting good practices on equipment purchasing, use and maintenance and health and safety. The objective of Pillar 5 of the GSP is to help soil laboratories to produce analytical results that can be compared, wherever the soil sample was analysed as long as metadata is provided.

Laboratory soil analyses refer to appropriate methods and procedures used to analyse the properties of soils. Generally, for the main soil characteristics, most of the laboratories use analytical methods based on the same chemical and physical principles, but many differences can be observed concerning the details of the analytical procedures themselves. The impact of the differences in analytical procedures for a given soil property on the final analytical results must be evaluated in order to estimate if the results coming from different laboratories could reasonably be compared. Thus, inter-laboratory comparisons (also called ring test or proficiency test) are needed to assess the performance of different testing laboratories which are measuring the same property on the same sample with the same method. Of course ideally, the results should be very similar.

Statistical analysis of the results is crucial to assess the quality and comparability of individual laboratory procedures as well as the overall lab performance. Results from different analysis procedures can be translated and combined where necessary to provide globally consistent data. The statistical evaluation consists of different procedures depending on the objective of the comparison, the number of laboratories involved, the type of material analysed, and the type of analyses. The statistical evaluation of the inter-laboratory comparison aims at providing information on the precision and accuracy of the soil analysis in order to assess laboratory performance. The precision refers to the degree of variation in repeated measurements of the same quantity of a specific parameter. It is the ability to provide the same result after repeating the analysis or to have little variation between two analytical results (EH&E, 2001; FAO, Suvannang and Hartmann, 2019). The accuracy refers to the

[148] http://www.fao.org/3/a-az922e.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

degree of correctness with which a measurement reflects the true value of the property being assessed (EH&E, 2001). In case of existing databases, conversion functions can be used to harmonize different analytical method values. A list of possible functions has been prepared by the Global Soil Map project (GlobalSoilMap, 2015) for organic carbon, pH, and texture. They note, however, that these functions generally are 'not-portable' to other pedo-climatic regions. Also, in the frame of the GS Soil project several algorithms have been collected (Feiden et al., 2011). Unfortunately, the web-portal of the project has not been maintained and updated, thus the deliverables contents are no longer accessible. A solution can be to make them available on the European Soil Data Centre (ESDAC) portal if all former project members agree.

Spectroscopy

A possible future way of transforming soil property data from one wet chemistry lab method and or lab to another is via spectrometry. This technique or method is increasingly used as a sufficient or constant quality and cheaper alternative to conventional wet chemistry lab analysis for a range of soil properties such as soil texture fractions, soil (in)organic carbon, total soil carbon, soil organic matter, pH, CEC, several nutrients, some bulk density/water retention related properties depending on the region and application.

Soil spectroscopy or spectrometry uses the interaction of light in the near-infrared and mid infrared part of the electromagnetic spectrum with soil to derive or estimate soil properties. The reflectance or absorbance of light in the infrared depends on the chemical properties of the soil (minerology and chemical (organic) bonds mostly) and accurate measurements of these soil spectra in the lab can be used to estimate soil properties. To enable this, a spectral library or calibration set is built from a large number of soil samples that are measured both spectrally and with wet chemistry (traditional lab) methods. This spectral library is then used to derive spectral models or 'correlations' between spectra and soil properties using often complex statistical and machine learning or AI models. This spectral model can then be used to estimate the soil properties of newly measured spectra. The (lab) method to be stored as metadata for this newly estimated soil property is then spectroscopy with the spectrometer, the spectral model used and the property/method calibrated to (this is possible in the Observation & Measurement standard, see chapter 3.3.5). The quality of the models depends on the quality and standardisation of the input data, both wet chemistry and spectra. Spectral models based on data from soil properties determined with the same wet chem method and lab will be of higher quality and able to better predict from new spectra than models that use data analysed with more methods and labs without transformation. The applicability for estimation of a new spectra for a certain soil property depends on the feature space of the soil properties (in soil types, minerology, range of values, presence of the wet chem soil property in the library) and the method and devise with which the new spectra are measured. Currently, standards for spectral measurements in both the VNIR and MIR range for a range of instruments, quality control procedures, spectral (meta)data storage and spectral calibration transfer research are under development in the GLOSOLAN initiative on Spectroscopy[149] (chapter 3.1.3.2Global Soil Partnership (GSP)) and the IEEE P4005 effort[150]. The

[149] http://www.fao.org/global-soil-partnership/glosolan/soil-analysis/dry-chemistry-spectroscopy/en/

[150] https://sagroups.ieee.org/4005/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

GLOSOLAN initiative furthermore is building a Global Soil Spectral Calibration Library and Estimation Service, which aims to build a truly global high quality spectral library for MIR to start with, later on to be extended to the visible near infrared (VNIR). This allows cheaper estimation of soil properties based on new spectra collected anywhere in the world, but also the possibility to create spectral models for different wet chem methods of the same soil property, which can estimate the soil properties as they would be upon analysis with those wet chem methods, based on spectra. In this way spectral libraries can be used to transform and harmonise soil property data from one wet chem lab method or lab to another and allow comparison of measured wet chem data on the same sample, thus providing a quality assessment and a possible transfer function.

Texture

As for texture, empirical equations to convert between the international system and the USDA/FAO system are available in literature: log-linear conversion (Nemes et al.1999), log-normal conversion (Shirazi and Boersma 1984; Buchan1989), multiple regression (Skaggs et al., 2001; Minasny and McBratney, 2001). A recent comparison is available in Takahashiet al. (2020)[151].

Several studies have been carried out to compare traditional methods, mainly Pipette and Hydrometer, to other more innovative methods such as X-ray Granulometer, Laser Diffraction and Morphology with the aim of replacing the former (particularly onerous in terms of time and laboratory work) with the latter, to reduce executive time and meanwhile gaining in accuracy. The ambitious perspective of rapidly measuring a wide range of particle size classes and using a small amount of soil sample, increased the use of Laser Diffraction in different laboratories. As a consequence, a huge amount of inherent literature was produced. Many studies not only developed possible conversion equations from Laser data to Pipette or Hydrometer ones but also determined the specific set up for both soil sample treatment and Laser analytic procedure (Di Stefano et al., 2012; ; Fisher et al., 2017; Bittelli et al., 2019; Al-Hashemi et al, 2021). Among the cited innovative systems only X-ray Granulometer is based on the sedimentation method, therefore its results are easily comparable with those provided by the traditional methods. Laser-Diffraction does not assume spherical clay particles as does the pipette method, therefore measuring fundamentally different fractions and is thus more difficult to transform. X-ray Granulometer automatically obtains a continuous curve in ten minutes, at the most. In Andrenelli et al., 2013 a particular set up of the X-ray Granulometer is illustrated (SediGraph 5120), tailored for soils, and is able to conspicuously reduce the time of sample preparation by combining the analysis by sedimentation with wet sieving for particles smaller and larger than 250 microns, respectively. This procedure succeeds in measuring up to 17 samples a day when an Autosampler (MasterTech Carousel) is available. Conversely, the standard procedure (pipette method) analyses sediments smaller than 63 microns and requires a time-consuming procedure to get the sample free of sand (Müller H. W. et al., 2009; Sporlein et al., 2004; Delaune et al., 1991). Actually, SediGraph is typically employed to determine grain size distribution of sediments in the environmental field (lacustrine, marine or fluvial) or powders (ceramics, metals, cosmetics, pigments) in different industries (Fossile E. et al., 2021; Yalamaç E., 2014).

[151] https://doi.org/10.1080/00380768.2020.1763143

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Another method used more and more for soil texture estimation (predominantly clay and loam fractions) is soil spectroscopy (modelled to pipette method fractions predominantly but not exclusively), discussed in the previous paragraph.

Soil Organic Carbon

In recent years, the reference method for soil organic carbon determination is the dry combustion method (ISO 10694). The Springer and Klee method results in comparable values, while the Walkey and Black method has to be corrected with empirical equations. A list of these is collected in GlobalSoilMap (2015). A recent work compared the Walkey and Black method, the Springer and Klee method with the dry combustion and a TOC analyser (Sleutel et al., 2007). The study found that very close relationships can be established between the OC content measured with modern dry-combustion methods and traditional wet-oxidation methods. Meersmans et al (2009) compared the classic and modified Walkley & Black method with the dry combustion method, obtaining a conversion factor of 1.47 which is higher than the usually applied value of 1.33.

Another method used more and more for soil organic carbon estimation is soil spectroscopy (modelled to both dry combustion predominantly but not exclusively), discussed in a previous paragraph.

Electrical Conductivity

The recent GSP initiative to map Salt-Affected Soils globally requires the harmonisation of the indicators chosen for the map, i.e. the electrical conductivity, ECe (dS/m), pH, the exchangeable sodium percent, ESP, and the sodium adsorption ratio. The reference electrical conductivity is measured on saturated paste extract (ECes) according to the GSS technical manual (Omuto et al., 2020). However, EC is often measured on other soil:water ratio extracts. Conversion equations are available in the literature (Landon, 1984; Hogg and Henry, 1984, Ozcan et al., 2006; Sonmez et al., 2008; Kargas et al., 2018). Some of these equations are reported in Omuto et al., 2020. As in the case of other transfer functions, these equations should always be recalibrated on local datasets.

Pedotransfer functions

In case of existing databases, a useful tool to further interpret and possibly harmonize data from soil inventories or values from different analytical methods is provided by pedotransfer functions (PTFs). PTFs are particularly used for soil physical and hydrological data to fill in gaps in measured data and to derive parameters and indicators that are difficult to measure. Bouma in 1989 introduced the term PTF for the first time to identify equations expressing relationships between soil properties. These equations, from simple linear regressions to the recent random forests, relate easily available soil information (e.g., soil texture, bulk density and organic carbon content) to soil properties and variables that are less available such as the parameters needed to model hydraulic, solute, thermal fluxes, and biogeochemical processes. Van Looy et al. (2017) provided a complete review on PTFs development, history and perspectives.

Data of significant quality and quantity of soil physical and hydraulic properties are often limited in soil databases because their direct measurements in the field or in labs are costly and time consuming.

Furthermore, such properties are commonly assessed with different methods, under different conditions, and based on different conceptual assumptions. For these reasons, *a posteriori* data harmonisation is often difficult if not impossible at all. Nevertheless, the knowledge of soil physical and hydraulic properties is required by relevant modelling applications (e.g., soil erosion, solutes and pollutants transport, soil-plant-atmosphere modelling) and for DSM applications (e.g., soil carbon stock, soil functions, soil-based ecosystem services).

The use of locally calibrated PTFs or of PTFs from literature validated and/or re-calibrated on local data allows for the estimation of soil physical and hydraulic properties suck as bulk density, saturated hydraulic conductivity and water retention, resulting in a harmonised set of estimates suitable for several applications.

Soil Bulk Density

Soil bulk density, i.e., the mass of soil of a known volume (g cm^{-3} or M gm^{-3}), affects basic hydraulic, electrical, mechanical, and thermal soil properties, playing a key role in the exchange of water, heat, and gas in the soil environment, in root growth and crop production. It is essential for weight-to-volume or area conversions. Soil bulk density is mainly affected by soil texture and structure, and by organic matter content, but it is also profoundly influenced by external conditions, i.e., land use/land cover, soil management, tillage operations, rainfall events, plant growth, cycle of freezing and thawing, and wetting and drying. Bulk density can be measured with direct or indirect methods, the former by collecting undisturbed soil cores of known volumes or by filling excavated soils with known volumes of filling materials (e.g., sand, water, seeds, marbles). The latter, indirect methods, use after *in situ* calibration, the attenuation of gamma radiation by matter, as the decrease intensity of radiation is a measure for the amount of matter in a known volume (Costa et al., 2013) or are combined penetrometer–moisture probes which provides an derived indication of bulk density (Vaz and Hopmans, 2001). Direct methods are by far the most used (Al-Shammari et al., 2018). As soil bulk density is a key state variable for the estimation of soil hydraulic properties, PTFs are often used for its estimation. Examples of reviews of the application of PTFs for bulk density estimation are given by Nasta et al. (2020), Abdelbaki (2016), De Vos et al. (2005), and Kaur et al. (2002). Different PTFs can be used to fill the data gap in existing databases, and the selection of PTFs should be based not only on the availability of harmonised input data but also on horizon depth and designation, soil texture, parent material and land use.

Saturated hydraulic conductivity

Saturated hydraulic conductivity (Ks) is essential to many soil science applications, but its availability in most soil databases is generally very limited. Notable exceptions are national and international datasets of hydrological properties assembled with the aim to provide scientists with the needed information while avoiding fragmentation and a general lack of standardization in Ks measurements. Examples are UNSODA (Leij et al., 1996; Nemes et al. 2001), HYPRES (Wösten et al., 1999), GRYZZLY (Haverkamp et al., 1998) and IGBP-DIS database (Tempel et al., 1996). More recently, the European HYdropedological Data Inventory (EU-HYDI) was established as an extension of HYPRES collecting additional data from European soils focusing on soil physical, chemical and hydrological properties

(Weynants et al., 2013). The EU-HYDI database is not distributed outside the participating institutions and external partners can access only derivatives for joint publications.

In most cases, though, data were collected from different sources and consequently different levels of standardization might occur, as Ks measurements were made under different field or laboratory conditions, over different soil volumes, resorting to different instruments.

In most applications, Ks is inferred via PTFs in order to reduce costs and save time. In a recent review of existing PTFs, Zhang and Schaap (2019) pointed out that there is still the need to adopt "consistent methods and protocols for measuring saturated hydraulic conductivity" and that "ensemble models for PTF development should be employed" using scale-consistent harmonised input data, keeping always in mind that in general the accuracy of Ks estimates is expected to be low.

Soil water retention

A third relevant soil physical property usually stored to some extent in databases is soil water retention, θ(h). Water retention data are generally the water content θ at specific water potential values h. More frequently, they are available for a limited number of potentials, typically at -1.5 MPa ("wilting point") and -0.033 MPa (field capacity) and at saturation. In order to derive the complete soil water retention curve additional points are required. According to Grigolon et al (2020) at least 7 points are necessary to ensure that the parameters of the continuous model describing the curve do not show statistically significant differences. Water potentials at different water contents can be measured in the field or in laboratory using different instruments with specific ranges of operability and accuracy. In the laboratory, the soil water retention curve is mostly determined using pressure chambers. Many authors nevertheless have pointed out that the assumption of full equilibrium of the sample at the designated pressure is not met at low water potentials (< -0.1 MPa) suggesting the use of vapor pressure methods (Bittelli and Flurry, 2009). In general, the combined use of different apparatus within different specific ranges of water potentials to measure water contents at different numbers of potentials, results in a lack of standardization across different databases or within the same database. Their standardization usually takes place by applying one of the continuous water retention models, most frequently the coupled van Genuchten model (van Genuchten, 1980). The use of a model allows for expressing the soil water retention curve by a limited number of parameters, which in turn allow estimating the soil water content at any potential. When any such standardized data are used, it should be considered that the equation generally does not fit perfectly to measured water contents.

3.6 Possible data storage environments

This section discusses the storage of data. There are different aspects to consider which will become apparent by answering the following questions:

- Where to store?
- How to store?
- How to share?

When it comes to technical aspects or possibilities in answering these questions there are basically no obstacles, meaning there is not one universal solution and depending on your requirements almost everything can be realised using available technology.

- Where to store?

Usually in the research and (non-)governmental domain, many parties are collaborating and many different users want to access the data. Data should therefore be stored in a network that guarantees sufficient access to producers and users, meaning a distributed (spatially dispersed) or federated (distributed and with a higher level of autonomy and heterogeneity[152]) network with nodes where servers contain the data that are accessible. In this way, the data owners keep full control on the data and are able to do the maintenance and regular update of the data.

Servers can reside with parties within the active working environment of the organisation or in the cloud where cloud providers maintain the servers. More and more cloud storage is being used because of its performance and scalability and because maintenance of the system is outsourced, a host organisation only needs to maintain the data itself, not the infrastructure. Both options are fine as long as adequate accessibility is provided. A decision may be based on the in-house availability of capacity (skills) and budget and legal aspects. Ongoing developments in IT but also within organisations, the use of specific standards and overall policy can also be a major factor in the decision where to store.

- How to store?

How to answer this question is to some extent depending on the way the producer creates the data. The choice on the design of your data structure (the data model or domain ontology) may also have a strong relation with the implementation in physical storage. This can range from relational databases (tables in RDBMs), to array data storage (NetCDF) to triple stores (linked data). Examples are Cloud databases, Cloud bucket storage, NoSQL, GraphDB. For each storage option, the management tooling is available, ranging from proprietary to open-source software packages also providing in many cases cloud solutions. Examples of physical data storage solutions are PostgreSQL, Oracle, MS Access for relational databases and MongoDB for NoSQL. So again, depending on historical use, a choice will depend on available capacity (skills) at the data owner and budget. And again organisational, topical and IT developments, use of specific standards and overall policy can play a major role in the decision where to store.

- How to share?

This is probably the most difficult but also the most crucial question to be answered, since it can influence how to answer previous questions. You need to know how access is granted (the license of the data, for more information see deliverable D6.2), the sensitivity of the data (when not to give access to data), who to authorise, how to authenticate, do you need accounting to monitor use or to charge users, etc.

In some cases, data are used to input them in user applications where they are processed. Here centralized, distributed and federated are different ways to process (distributed) data. There is a lot of confusion about these terms as they can easily mean different things for different purposes. So, for

[152] https://en.wikipedia.org/wiki/Federated_architecture

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

databases, architectures, networks and even governments, these terms are used as adjectives and as such give different meanings to the definitions.

But there are also similarities, Federated means in all cases working together, to collaborate. Centralized and distributed have a spatial concept, that is, in one place or in different places (can be spatial but also within one organisational unit). See also the descriptions in the Where to store section on the previous page.

In general, OGC or W3C standards are used to set up API's or webservices to actually share data. For example the INSPIRE geoportal contains metadata with all the links to the view services (WMS) and download services of the shared data using the web mapping services API's from OGC. For EJP SOIL we choose a distributed or federated model, where all countries store and keep their data but share the metadata and provide OGC API's through the network.

REST and SOAP Web-Services:

Web services are a set of standards and open protocols used to share data between two or multiple separate applications or systems. The software applications can be written in different programming languages (java, c#, php, ...) and running on various platforms (windows, Linux, ...).

To connect a service, developers don't need to write code from scratch, they can integrate a ready-made service usually after choosing a suitable approach like SOAP or REST.

The first one is SOAP, it stands for "Simple Object Access Protocol" and has been around since the late 1990s. It's a messaging protocol for interchanging data in a decentralized and distributed environment.

The second tool is REST, it stands for "Representational State Transfer" and it is the newcomer. It seeks to fix the problems with SOAP and provide a truly simple method of accessing web services. REST allows API providers to deliver data in multiple formats such as plain text, HTML, XML and JSON. It's currently the most popular choice of developers to build public web services.

When designing their API's, developers must answer the age-old question: Should we use SOAP or REST? To help choose between them, below a comparison table of REST and SOAP that highlights the main differences between the two approaches is given:

	SOAP	**REST**
Concept	Standardized protocol with pre-defined rules to follow	Architectural style with recommendations
Approach	"Remote method invocation" (e.g : getSamples())	Data available as resources e.g : « samples »
State	Stateless by default, also supports stateful	Stateless (no server-side sessions)
Caching	API calls cannot be cached	API calls can be cached
Performance	Requires more bandwidth and computing power.	Requires fewer resources
Security	WS-Security protocol	SSL and Https
Messages Format	Only XML	XML, Json, html…
Verbosity	Verbose (XML, WSDL, Message composed of: Envelope, Header, Body)	Consise (smaller messages formats)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Supported Protocols	Http, Smtp, Udp...	Only Http
Advantages	Standardized Security	Flexibility Scalability Better performance Browser-friendliness Smaller learning curve
Disadvantages	Less flexibility More complexity Poorer performance	Less security
Usage	Financial services Payment gateways Telecommunication services	Mobile services, Public API's for web, Social networks...
	SOAP	**REST**
Example	APIs generally developed in-house (eg : banks)	SensorThings, AWS, Google Maps, Tesla, Flickr, Twitter, Facebook..

To summarise, REST is device independent and a client consuming REST API can be anything like Mobile devices, Notebooks, Tvs etc. With the popularity of the cloud environment, applications are slowly moving to cloud based systems such as Amazon AWS, Azure. These systems are built on and are exposing REST API's. Hence it is a good move to build applications on top of the REST API.

To summarize, what you need to decide for your technical environment is based on answers to the above questions. It comes down to knowing the requirements for data usage by producers and users. In general, one and the same choice does not have to be made for all components of the data storage environment, especially in a distributed environment. Importantly, there are many ways and converting tools available to map, translate or transform data that is exchanged in a standard format. That data can be imported into your own environment for use. Keep in mind that the data must be semantically harmonised to facilitate easier technical conversion. Once you have made these choices, it is easier to decide which storage and sharing tools are most suitable for your setup and which can facilitate your design and use.

3.7 Input of data

Ontologies, codelists, domain models, cloud storage and metadata standards are terminology and are concepts that may sound alien to many soil scientists, including the scientists or field experts that perform actual soil sampling, are gathering information to make a map or just want to use existing soil data to run for instance their hydrological model. At the same time, these concepts and data science are needed to handle an increasingly complex soil data universe and to make it possible for these soil scientists to find the data they need at any given time on any given topic for any given aim. This means that soil and data scientists need to find a middle ground where soil scientists are enticed to adhere to standards and structure their data collection using easy to operate tools. And where these tools allow data scientists to work with an error-free or error-low input result that enables them to focus on the data modelling, storing, serving work instead of tracing input errors one by one.

The two worlds meet when new soil data is collected and when older, existing databases need to be included in or mapped to a new data model. Several tools exist that facilitate this, although much depends on the case-by-case design and use of these tools. When disregarding the traditional printed field form (very error-prone due to handwriting issues, rain, possibility to deviate from the classification rules), a first option is the use of restricted Excel sheets. These sheets can be restricted by applying data validation rules to allowable content in cells. This can include minimum and maximum values, data type (integer, string, decimal), restricted dropdown list according to the codelists or standards that are applicable for a campaign or property.

A second option is the use of customisable proprietary or open-source software such as ArcGIS Field or QGIS. This can be customised to only allow valid observations according to the relevant standard (e.g. a soil classification method). A third option is software like ODK[153] which allows the user to set up questions and possible answers in an Excel form and then expose that to an app. This app can be used in the field during data collection, thus limiting the amount of possible answers and allowing to tune the survey list to the campaign at hand. The result can be exported from the server as an Excel file for storage in the appropriate data storage environment or database system.

Another way of inputting data into a data storage system is ingest it from already digitally available (online) resources. This can be an API of another (soil) information system, either in the soil domain or beyond. In this case, it will most often follow OGC standards for exchange of geospatial data or other web standards. When data are to be ingested or used (distributed or federated approach) that is not provided through an API but is for instance an older database on a local computer or disk, often a mapping is needed between data models. The data model of the source and of the target database or environment should be known, including the ontology and codelists/vocabs that describe the grammar, definitions and properties of the data. Well described metadata are absolutely crucial for this to succeed. Standardisation or even harmonisation of data sources becomes very difficult if not impossible without introducing large uncertainties to the combined datasets. Although a soil scientist should be involved when making a mapping between data models or datasets, often a data scientist is needed as well when datasets are more complicated that one or two Excel sheets. This ensures reproducibility through code and can prevent errors in copy pasting in Excel. At the same time, a soil scientist is needed to provide the much-needed domain knowledge for a mapping to be done correctly.

3.8 Serving data

Two examples are described below. One from France with the use of webservices through SensorThings API and exploring web semantic technologies, the other from Flanders, which presents how they harmonize and disseminate soil data with Geoserver (for data) and Geonetwork (for metadata).

[153] https://opendatakit.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Recently the new EU directive 'Open Data Directive' 2019/1024[154] was published. This indicates that the harmonisation process of data creates datasets that are geospatial high value (Annex1). And it will support public API development as well as bulk download and access to soil data.

3.8.1 Example on how to disseminate soil data from the French Soil Information System

The strategy for disseminating open soil data from the French Soil Information System was first based on Geonetwork[155] in order to fulfil the INSPIRE Directive. Geonetwork enables to describe the geospatial metadata using the INSPIRE metadata format but also to give access to spatial data through webservices for visualisation (WMS) and download (WFS). However, these types of webservices are not adapted to measured data, especially those measured with sensors (Kotsev et al., 2018).

In 2018, INRA developed a portal using the open-source Dataverse[156] software developed by the Institute for Quantitative Social Science of the Harvard University. The INRA portal is available for all researchers of the Institute. It provides a quick and easy way of disseminating research data of any types of files (csv, shapefiles, etc.). Although this tool facilitates the data dissemination for researchers, interoperability is not really implemented yet.

In two French research projects (project Ademe-BRGM FGU III, project ANR Data4C+), interoperability of several databases was scheduled and for this, we choose to use the SensorThings API version 1.1 (Lattelais et al., 2021a). This API is based on the OGC standard Observation and Measurement (OGC, 2011) and on IoT (Internet of Things) technologies providing a RESTful interface. We follow the recommendation of Kotsev et al. (2018) who proposed to use the SensorThings API for INSPIRE implementation for measurement data whatever the theme.

To implement the SensorThings API, we use the FROST server implementation, an open-source implementation developed by the German research institute Fraunhofer IOSB. We develop also specific open-source tools:

- SensorMap for the mapping between our relational database in Postgres[157], which stores the soil data, and the SensorThings data model. It works also with CSV files.
- SensorBoard, a dashboard for SensorThings data visualisation

During the process of using the SensorThings API on our data and also to be compliant with the INSPIRE Directive, it is necessary to give the list of the observed properties, measurement methods and units of measurement. To do this, we use a process based on semantic web technologies and described in Lattelais et al. (2021b). Our objective is to deploy this information as RDF triples on a SPARQL endpoint in order to propose URIs for the different deployed objects with their interrelations. The first step is to create a template in graphML format using the desktop application yEd graph editor. This template has three parts:

[154] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information
[155] https://geonetwork-opensource.org/
[156] https://dataverse.org/
[157] https://www.postgresql.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

- The annotation graph which describes the different concepts, their properties or relations and how data are mapped,
- the SQL queries to retrieve the data by the mapping from the PostgreSQL relational database,
- the URI definition for the different concepts described in the annotation graph.

This template is then used by the pipeline Coby, developed for the AnaEE France project, which:

- applies the SQL query on the relational database,
- transforms the data in RDF triples following the annotation graph,
- publishes the RDF triples on the SPARQL endpoint (based on a Blazegraph triplestore).

This work is ongoing and for the moment, we focus on the implementation of the whole process using the example of soil organic carbon data. We now need to apply it to other soil data measurements, to enhance the data interoperability, and to establish relations with controlled vocabularies or existing ontologies.

However, our tools are developed for a use with relational databases. Researchers do not always have access to well-structured relational databases.

We have chosen to develop an interoperability based on web services because their development is fast and easy and makes it possible to respond within a reasonable time to the request for soil data interoperability. Semantic web technologies take a longer time to develop because they require controlled vocabularies, ontologies which suppose an agreement on the concepts among researchers, and works on semantic annotations. But they are more durable and more adaptable than web services. Thus, we plan to apply the semantic web pipeline on the soil data. The difficulty is to determine which ontology to use. We are exploring existing ones, especially the AnaEE ontology which could be used with development of specific extension.

3.8.2 Practical example INSPIRE harmonisation for Soil in Flanders

Soil data in Flanders are stored in different schemas, tables and views in a PostgreSQL (with PostGIS[158] extension) database. The publication of the actual datasets/map layers are managed through a task manager function in GeoServer[159]. The soil data can be viewed in the online 'DOV-verkenner' or can be accessed by web services[160].

For the INSPIRE harmonisation we work a little different: we put all the data together in a small number of materialized views. These views are a combination of a table and a regular view in which multiple tables are joined together. They are independent from other data. This way we can much easier assign the data to the Soil-schema. For larger sets of data, this is indispensable from a performance point of

[158] https://postgis.net/

[159] http://geoserver.org/

[160] https://www.dov.vlaanderen.be/portaal/?module=verkenner&bm=34c4e592-149f-4449-9c9b-0a477d115391

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

view. The drawback of this is that for regularly updated data the materialized view needs to be recreated (which can easily be done by a scheduled task on monthly, weekly or daily basis).

Once the materialized view has been created, we start mapping to the represented featuretype of the INSPIRE Soil-schema. We choose to work with app-schema[161], an extension to GeoServer, which can cope easily with complex features. These are features that contain nested properties and work on multiple XSD-schemas. App-schema is actually nothing more than one or more XML-files referring to the materialized view(s) and the featuretypes from the INSPIRE XSD-schema with their properties. The from-to mapping (data to schema) could be done in a simple (XML-) editor or by using a software that is built for transformations of complex data sets, e.g. Hale Studio[162] from WeTransform (open source) (Figure 3.15).

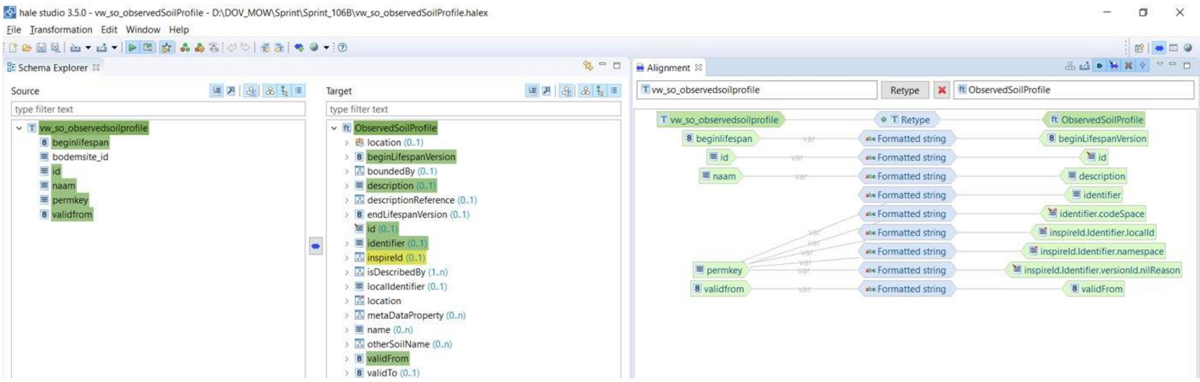


Figure 3.15 Example of the use of HALE Studio. On the left side we have inserted the materialized view (our data as source), in the middle the Soil schema was loaded (as target) and on the right side the links between the attributes of both are visually demonstrated in the alignment.

When the app-schema file(s) has been created correctly, we can implement it in GeoServer. We use the GeoServer Task Manager tool to configure the desired tasks and batches, starting with the app-schema publication task. Eventually we do the same as in a regular mapping publication adding a style from the portrayal in INSPIRE Soil data specification[163]. A metadata-task from the Task Manager tool will finish our publication. Adding the right keywords to the metadata of our published layer, we can harvest the harmonised layer into the GeoNetwork catalog application.

Besides these metadata, the entire process results in a view service (WMS) and a download service (WFS). The latter produces a XML/GML file for reading and visualizing the data. A downloadable result of one ObservedSoilProfile (with all its joined attributes) can be found here[164].

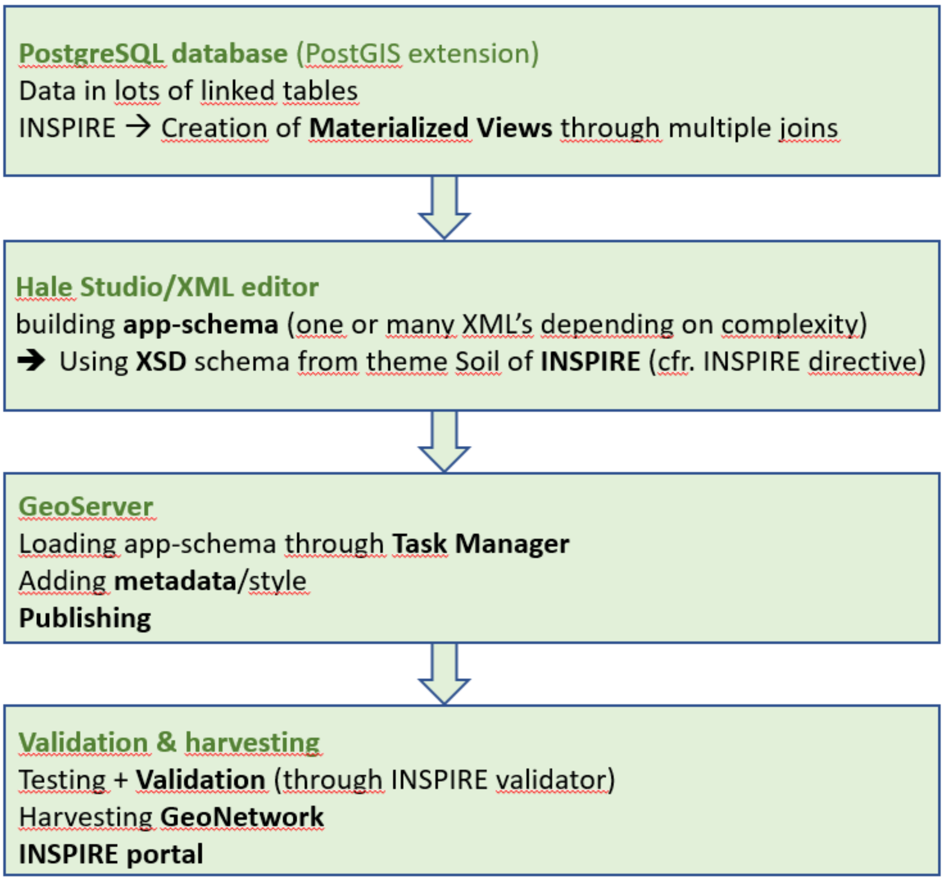
161 https://docs.geoserver.org/latest/en/user/data/app-schema/index.html
162 https://www.wetransform.to/products/halestudio/
163 https://inspire.ec.europa.eu/id/document/tg/so
164 https://www.dov.vlaanderen.be/geoserver-inspire/so/ObservedSoilProfile/wfs?Service=WFS&Version=2.0.0&Request=GetFeature&Typename=so:ObservedSoilProfile&count=1&outputFormat=application%2Fgml%2Bxml%3B%20version%3D3.2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

The result of the harmonisation can be found on the INSPIRE portal[165].

Validation can be performed through the INSPIRE validator[166], however only some general GML-tests on the used (soil)schemas can be executed as there is up to now no specific test on the Soil theme available.

All software and methods used are open source (PostgreSQL[167]/PostGIS, Hale Studio[168], app-schema[169], GeoServer[170], Task Manager[171], GeoNetwork[172]) and can be implemented by anyone without additional costs. The combination of systems that is most appropriate or suitable to use depends on the database on the data owner side and the INSPIRE schema and can therefore be different for different data owners and situations.



165 https://inspire-geoportal.ec.europa.eu/download_details.html?view=downloadDetails&resourceId=%2FINSPIRE-f0c91711-ece0-11e8-a08e-52540023a883_20210123-164202%2Fservices%2F1%2FPullResults%2F361-380%2Fdatasets%2F18&expandedSection=metadata
166 https://inspire.ec.europa.eu/validator/
167 https://www.postgresql.org/
168 https://www.wetransform.to/products/halestudio/
169 https://docs.geoserver.org/latest/en/user/data/app-schema/index.html
170 http://geoserver.org/
171 https://docs.geoserver.org/latest/en/user/community/taskmanager/index.html
172 https://geonetwork-opensource.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Figure 3.16 Overview of the INSPIRE harmonisation process

There are some unsolved issues in the INSPIRE implementation of soil data in Flanders:

- In INSPIRE Soil theme, the most difficult featuretype (by far) is the ObservedSoilProfile because it is linked to many other featuretypes like SoilPlot, SoilHorizon, SoilLayer, SoilSite, ProfileElement and items like classification (WRB or other). The many observations to each of the mentioned featuretypes can only be linked directly in the draft schema for soil and not in the official soil-scheme (4.0). It is necessary that the draft schema becomes the official one. It is necessary that the draft schema becomes the official one.
- Where is soil erosion fitting in in INSPIRE (and all related maps)? Where are guidelines for soil erosion in INSPIRE?
- How to implement the raster data into INSPIRE coverages?
- Some codelists for soil are missing or are incomplete
  - A lot of empty code lists that cannot be extended
    - https://inspire.ec.europa.eu/codelist/WRBQualifierPlaceValue (https://inspire.ec.europa.eu/codelist/WRBQualifierValue
    - https://inspire.ec.europa.eu/codelist/WRBSpecifierValue
    - https://inspire.ec.europa.eu/codelist/FAOPrimeValue
    - https://inspire.ec.europa.eu/codelist/FAOHorizonSubordinateValue
    - https://inspire.ec.europa.eu/codelist/FAOHorizonMasterValue
  - Codelists that are not complete: like the WRB code lists are only for WRB 2007 and not for WRB 2014 and also the parameter codelists are far of complete:
    - https://inspire.ec.europa.eu/codelist/WRBReferenceSoilGroupValue
    - https://inspire.ec.europa.eu/codelist/SoilProfileParameterNameValue
    - https://inspire.ec.europa.eu/codelist/SoilDerivedObjectParameterNameValue
    - https://inspire.ec.europa.eu/codelist/ProfileElementParameterNameValue
    - https://inspire.ec.europa.eu/codelist/SoilSiteParameterNameValue
- We are missing definitions on vocabulary and ObservedProperty in Flanders (work in progress)
  - Is there a common registry that can be used for the procedures (methods) used for the observed properties (xlink:href)? We are missing an URL for parameters for the observedproperty. For the theme water for example, this is available but not for soil: for example http://dd.eionet.europa.eu/datasets/latest/NiD/tables/NiD_GW_Conc Are there already existing URLs on European level for soil parameters that can be used?
- Soilbody: there is no place to put the SOILUNIT of the WRB classification of a soil polygon when you have no DerivedSoilProfile. So what has to be done when you have a soil map without

DerivedSoilProfile but with a SOILUNIT according to the WRB classification? (solution for Belgium: SOILUNIT in gml.description but this solution is not a good solution because the main information (SOILUNIT) is somewhere hidden in an extra field)

It is not clear how to proceed when you have a question about the soil scheme of INSPIRE, when you have a request to change or extend some codelists or when you ask for the draft scheme to become official. Official contact points for INSPIRE SOIL and the procedures how to proceed for questions, changes or extensions are needed.

The issues identified by Flanders are taken up as action items in EJP SOIL WP6 and show that it is at present (early 2021) not straightforward to implement INSPIRE Soil for partners and member states. One of the aims of EJP SOIL T6.1 is to solve as many of these issues as possible to enable a distributed soil data infrastructure based on the INSPIRE directive including the Soil theme in Annex III.

3.9 Choices within EJP SOIL

In this chapter the ongoing initiatives, background, basic principles and choices are addressed for setting up a soil information system while using available standards for data storage, exchange and harmonisation of soil data. Following these structures, it will be easier to organise, store, use and exchange soil data for research, policy and other applications. When setting up a SIS, there is not one single best way to do it because every situation has its own requirements and therefore choices in architecture, data standards etc. However, one choice is crucial to get started. The basis for any setup is always the semantic model (or conceptual model, or information model that are synonyms and all mean the same). A semantic model is independent of the technical implementation. The level of detail required is determined by its application and thus determines the scope or extent of the domain that is described. A semantic model formalizes and synthesizes human knowledge in that particular domain and is currently referred to as a domain ontology in most cases. An ontology, still implementation independent, can be described in different ways, such as natural language, class diagrams in UML or knowledge graphs. Only during the implementation the (technical) choices are made. The terminology, principles, questions and examples given in this chapter can assist in choosing the most appropriate setup in a given situation and offers good practices that apply in all cases. These include a good documentation of metadata, adherence to existing standards such as INSPIRE, OGC, ISO, Dublin Core etc., and making data findable, accessible, interoperable and reusable (FAIR).

Within the EJP Soil programme, we aim to set up a distributed soil information system that adheres to and uses the INSPIRE principles of the directive, the metadata specifications and data specifications for soil (Annex III) as a base ontology, that can be extended for broader use or specific applications. This means that we choose that data remain at and is curated and updated by its owner (institute) and can be exchanged by a common infrastructure using the INSPIRE soil domain model and appropriate technology. The examples and the overview of harmonisation possibilities show that there is still quite some work to be done before harmonised soil data can be exchanged effortlessly by partners and or member states and the EC/ESDAC/EU Soil Observatory following the INSPIRE model. This can be independent from the way partners choose to organise their data. At present we can differentiate two main possible directions or ways to come to a distributed soil information system within EJP:

1) the quick way using webservices without a fully implemented (INSPIRE) soil ontology that is agreed upon by most partners (for instance the example by France (see 3.8.1 and GLOSIS version 1). This will

allow exchange of data but in cases will need transformation to be easily readable or understandable by other systems or partners that use other encodings.

2) developing an easy implementable INSPIRE based soil ontology (that means to have a consensus about the data model) by extending the INSPIRE core that is available and choosing a technology for implementation such as webservices, APIs or the semantic web technology (eg. GeoSPARQL).

What matters here is that data connectivity between data services must be guaranteed without loss of data and relationships (ie information), regardless of the chosen implementation. This will allow easy and standardised exchange of soil data in Europe and allows partner institutions and others to comply with the INSPIRE Directive (Annex III), but will most likely need a lot of work and discussion. The example of Flanders (3.8.2) shows this.

The latter is the most extensive and time consuming one but also the most realistic for the future. This chapter provides the basis for discussion within EJP SOIL WP6 on these possibilities. Currently ongoing activities in WP6 are aimed at resolving as much as possible the present impediments for full and easy implementation of INSPIRE Soil by partner institutions and member states. This is geared towards at some point in time arriving at the latter option of a full-fledged standardised decentralised soil information system for Europe that allows the use of (at least semantically) harmonised soil data for many different applications. Whether or not that is feasible within the timeframe and possibilities of EJP SOIL, or if we will work towards that but need to choose the first option of a more 'light-weight' implementation of a decentralised soil information system for EJP will be decided in close cooperation with EJP WP6 partners during regular intermediate web-meetings. Final specifications are therefore delivered in D6.8 at the end of the project (M59) and can be shared with ESDAC/EUSO in the meantime.

The next steps for this work will be to write a concept note style work plan that outlines:

- Activities required by partner institutions or countries that would like to take part in this soil information infrastructure, set up by EJP SOIL with the aim of facilitating soil data exchange according to INSPIRE in Europe.
- Activities required by the WP6 team as a whole to enable the creation of the infrastructure and resolve open INSPIRE Soil ontology implentation challenges, assist with tools for mapping and or format transformations, align with T5.5 on capacity building.
- The process and principles towards writing D6.4 which includes more details on the specifications of the system.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

4 Sampling theory for mapping and monitoring purposes

Authors: K. Teuling, B. Kempen, M. Knotters, N. Saby, D. Brus, R. Vašát, F. van Egmond, A. Bispo

This chapter focuses on the statistical aspects of sampling design for the monitoring and mapping of soils which imply the selection of the statistical sampling approach and the statistical inference. It starts with an overview of recent and relevant EU projects that have a significant component on soil sampling (strategies) to provide context and the state of the art on this topic in current EU projects. This is followed by an introduction to soil sampling design that can serve as guide to choose the most appropriate design depending on the aim of the sampling. But since many soil sampling surveys do not start from scratch or need to be maintained and updated over time it is not always possible to choose solely based on the aim. How to address these situations is described in chapter 4.3 including a direction on how to combine two surveys or monitoring systems. This will be elaborated further in a subsequent EJP SOIL deliverable D6.3, due in summer 2021. Other aspects that are important for soil sampling such as the sampling protocol and metadata collection are only briefly described in 4.4 and will also be elaborated further in deliverable D6.3.

Chapter 3.6 addresses ways to store data about the sampling itself and its surroundings in the field, while chapter 3.4 addresses the importance of harmonised lab methods and data. Chapter 5 addresses the creation of maps once in situ point data is collected by means of sampling.

4.1 Relevant previous or ongoing EU projects

4.1.1 ENVASSO

The ENVASSO project[173] aimed to deliver a framework to monitor, at a continental scale, the condition of different types of soil that are subject to a range of threats. Its objective was to describe a common framework to enable a progressive harmonisation of current and future soil monitoring activities in EU Member States. Eight threats to soil (erosion, organic matter decline, contamination, compaction, salinisation, decline in biodiversity, soil sealing, and landslides & flooding) are identified in the Thematic Strategy for Soil Protection in Europe. All of these have been addressed in ENVASSO except for flooding. In addition, desertification has been included as an additional 9th threat.

The outputs from ENVASSO are a series of technical reports documenting criteria and indicators for the characterisation of soil, provided an inventory of monitoring systems, a database system suitable for data capture, the procedures and protocols for inventory and monitoring , and the results of evaluating prototype procedures and protocols in a number of pilot areas in Europe. It reports detailed recommendations for a future European Soil Monitoring System.

[173] https://esdac.jrc.ec.europa.eu/projects/envasso

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

ENVASSO has developed a system to harmonise existing, mostly national soil monitoring networks and databases, to form a European-wide reference that can assess current and future soil status and support the sustainable management of soil resources.

4.1.2 LUCAS and Landmark

4.1.2.1 LUCAS-Topsoil sampling design

In the Landmark[174] project the LUCAS-Topsoil sampling design was evaluated from a statistical point of view. LUCAS-Topsoil is a 10% subsample of LUCAS, a spatial sample for monitoring land cover and land use in Europe that started in 2001. The LUCAS design was completely redesigned in 2018. LUCAS-Topsoil is a subsample of the LUCAS sample designed in 2009. We will first describe the sampling design of LUCAS-2009.

4.1.2.2 LUCAS: subsample of square grid

LUCAS-2009 (hereafter shortly LUCAS) is a stratified subsample of the 2 km x 2 km grid covering 25 member states of EU (EU27 except Bulgaria, and Romania), using seven land cover classes as strata (see Jacques and Gallego: The LUCAS 2006 project – A new methodology, and Gallego and Delincé, 2010). The strata are determined at the 2 km x 2 km grid nodes using orthophotographs or satellite images (Eurostat 2012). Sampling rates differ between the strata, see Table 1 in Jacques and Gallego. The total number of georeferenced points in LUCAS is 19,967. Within strata points are selected in such a way that spatial clustering is avoided. This has implications for the estimation of the variance. The inclusion probabilities of the LUCAS points are perfectly known, and as a consequence unbiased design-based estimation of means and of reporting units is feasible. However, no unbiased estimator of the variance is available, due to the sampling design used in the first phase (systematic random sampling) and the second phase (cluster random sampling in a way that clusters are separated in geographical space). The variance can only be approximated.

4.1.2.3 LUCAS-Topsoil 2009-2015: subsample of LUCAS

LUCAS-Topsoil is a 10% subsample of LUCAS. The subsampling design is quite poorly described in Tóth et al. (2013). Four terrain attributes (altitude, slope, curvature and aspect, all derived from SRTM digital elevation model) and land use (derived from CLC2000) were used to construct about 2000 strata. These strata are obtained by overlaying five maps with (univariate) strata. Each terrain attribute is used to compute eight strata with an equal number of pixels (this is done by estimating quantiles of the frequency distribution). The final map obtained by overlaying consists of 30,795 polygons. The LUCAS points within a polygon are used as primary sampling units in two-stage random sampling.

Sample sizes were set for combinations of land use and country. The LUCAS points within a polygon were randomly grouped into triplets. From each triplet only one point is selected and sampled for observation. If the first point of a triplet is inaccessible, or for whatever reason unsuitable for soil

---

[174] http://landmark2020.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

sampling, another point is sampled et cetera. The soil surveyor is free to determine the order in which the points are visited in the field, which saves time needed for fieldwork, but compromises the computation of inclusion probabilities. If the number of triplets is smaller than the required sample size, more points are selected from the polygons containing more than six LUCAS points. Further details about how this is precisely done are not given. If the number of triplets exceeds the required sample size, those triplets are selected from the polygons of that specific land use that contain the largest number of LUCAS points. Note that the selection of primary sampling units is not random: the selection is targeted at the largest polygons.

4.1.2.4 Is design-based estimation of means feasible with LUCAS-Topsoil?

Tóth et al. (2013) motivate the choice for random sampling by the superiority of the design-based sampling approach for mapping purposes: "Since soil mapping, even top-soil mapping, is best performed if design-based, a multi-stage stratified approach (McKenzie et al. 2008) was chosen ". Based on the available information, we think that the LUCAS-Topsoil sample is a stratified two-stage "partly-random" (or fully non-random) subsample of LUCAS-2009. We conclude that the inclusion probabilities of the LUCAS-Topsoil points are not traceable, and besides there are points with inclusion probability 0. For that simple reason design-based statistical inference, qualified as superior by Tóth et al. (2013) (of which we fully agree) is not feasible.

In their paper and report Tóth et al. (2013) compute sample means and sample standard deviations for land use/land cover categories within the nine climate zones, for instance mean and standard deviation of SOC concentration for annual croplands and for permanent croplands within the nine climate zones. The sample means are used as *estimates* of the subpopulation means. Differences in inclusion probabilities of the LUCAS points due to different sampling rates in the LUCAS strata, are not accounted for. In Fig. 6.38 of the report, showing mean C:N ratios for four land cover classes within four climate zones, also error bars are shown, representing the standard error of the estimated means. It is not explained how these standard errors were estimated. Possibly these are computed by the sample standard deviation divided by the square root of the sample size.

In summary, we conclude that design-based estimation of the (sub)population mean and its variance from the LUCAS-Topsoil is not feasible. The only option is model-dependent inference using, for instance, a geostatistical model (kriging). The estimated means and their estimated standard errors as reported by Tóth et al. (2013) are model-dependent estimates, that follow from the assumption of independent data (i.e. a pure nugget variogram). The question is how realistic this assumption is. In any case such modelling assumptions can best be avoided. It is highly recommendable to select the sampling locations for monitoring the soil by a well-defined probability sampling design, so that means and totals of reporting units can be estimated either model-free by design-based inference, or by model-assisted inference.

4.1.2.5 National Soil monitoring network in Europe

Another part of the LANDMARK project looked into soil monitoring networks in Europe and investigated the description of existing soil monitoring networks (SMN) in Europe with respect to soil functions (van Leeuwen et al., 2017). This study concluded that:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

- Current SMNs form an unbalanced dataset, in which predominantly chemical soil parameters are included, but soil biological and physical attributes severely underrepresented.
- A wide range of different methods is being used in the different SMNs for measuring attributes.
- An intensive programme of harmonisation of the different methods used is necessary to permit valid spatial and temporal comparisons both within and between countries.

4.1.3 EBONE project

The EBONE (European Biodiversity Observation Network) is a project that focused on the development of a cost effective system of biodiversity data collection at regional, national and European levels between 2008 and 2012. Even though the design has since been updated and the focus was not specifically on measuring soil parameters, the sampling designs that were developed can still be considered as relevant examples of EU-scale monitoring possibilities.

The proposed sampling design is based on probability sampling, with the goal of providing unbiased and valid estimations of status and trend parameters and their standard errors without the use of models. Brus et al. (2011) propose two options of sampling designs;

- A stratified random subsampling of the LUCAS sample
- A stratified random sampling from geographical substrata within environmental strata.

In both cases the environmental strata (Metzger et al., 2005) are used as strata, which are relevant because of their ties with biodiversity. As mentioned in 4.1.1, LUCAS is a grid-based spatial sample. Within the first sampling design it is proposed to make use of these LUCAS locations and let them spatially coincide with the EBONE sampling locations. This can be attractive from an operational and financial point of view. However, because of the sheer size of LUCAS (250,000 points) a subset is needed. This greatly increases the complexity of the resulting sampling design, leading to complicated estimators and statistical inference. Moreover, it results in biased estimators of precision of the monitoring results.

The second design is solely based on the subdivision of strata that were based on the environmental strata. This geographical subdivision results in a good spatial coverage and also an unbiased estimate of the spatial mean. Also, the statistical inference is much less complex.

4.1.4 Soils4Africa

Soils4Africa[1] is a recent project that started in 2020 that aims to deliver continent-wide soil information, as does LUCAS Soil for Europe, relevant for agricultural intensification in Africa. For this purpose, soil samples will be collected from 20,000 sampling sites across agricultural land in Africa. The samples will be analysed in one laboratory in South Africa for relevant soil properties. From these, a set of soil quality indicators and functions will be derived relevant for sustainable intensification of agriculture. These data will provide a continental baseline of the status and prevalence of soil

[1] https://www.soils4africa-h2020.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

properties, quality indicators and functions that can be used for future monitoring following the same design. The data will be served through an online soil portal hosted in Africa and several use cases will be developed to showcase the use of the acquired data. At present (March 2021) a map of agricultural land has been derived and the design of the soil sampling scheme is ongoing. Results will be published on the website.

4.2 Designing sampling schemes for survey, mapping and monitoring

Developing a sampling design, for instance for soil sampling, is complex. It requires many design decisions that depend on the purpose of the sampling campaign, required results and resources that are available. The aim of this section is to offer guidance in designing sampling schemes and highlighting key design choices. To this end, we focus here on three types of sampling:

- sampling for estimating means, totals and (areal) fractions for defined geographic areas of interest. These target quantities are often referred to as 'global quantities'. One does not estimate these for a specific location but for larger geographic areas of interest such as for instance countries (e.g. national reporting on total carbon stocks), intervention areas, management zones, all arable land in a country, areas covered with forest, etc.
- sampling for mapping
- sampling for monitoring; i.e. for estimating global quantities in space at different times and how these quantities change over time.

When designing a sample scheme a key decision is the choice for a design-based or model-based approach for sampling and statistical inference. Design-based inference implies that the sampling units are selected randomly and are not based on convenience or prior information (purposive sampling). In general, this means that a computer-assisted approach is used to determine the coordinates of the locations using a random number generated based on a specified design (Brus, 2019). The strength of a design-based sampling design is that the probability of selecting a sampling unit from the target universe (i.e. the *inclusion probability*) is known and can be used for statistical inference. This type of design is best suited for estimating global quantities such as means and totals for the entire target areas or several subareas (domains of interest; see below) and changes of these quantities over time, because it provides valid and unbiased estimates of these quantities as well as the confidence intervals of the estimates.

In a model-based approach for sampling and statistical inference, the inference is based on a stochastic model of spatial (and temporal) variation. Probability sampling, i.e. random sampling so that the inclusion probabilities are known, is not strictly needed. In case geostatistical models are used, the weights of the data in case kriging is used to estimate spatial correlation (to predict at an unsampled location or to estimate the mean of an area) are determined by the covariances between the observations, which are based on the coordinates of the sampling locations. This type of design is better suited for mapping and makes use of spatial correlation structures of the target variable or cross-correlations with co-variables (De Gruijter et al., 2006).

De Gruijter et al. (2006) explain that the choice between design-based and model-based inference has major consequences for sampling. Design-based inference requires some form of probability sampling, while model-based inference commonly builds on purposive sampling (non-probability sampling). Furthermore, model-based inference requires a model of spatial variation. Development of such a model will usually require an extra sampling effort. Inference about a model mean often requires a considerably larger sample than inference about a spatial, temporal or spatio-temporal mean with the same level of precision. For more information about design-based versus model-based approaches for soil sampling we refer the reader to seminal work of De Gruijter and Ter Braak (1990) and more recent work by Brus (2021).

Changing conditions may affect the observations made during the sampling. As an example, seasonal variations in soil pH are well known and can be greater than either the spatial variation or laboratory error. As a consequence, if we are interested in spatial means or spatial patterns at a certain moment or in a certain period, rather than in variation in time, it is recommended to apply important restrictions to the sampling period. For instance, in case of monitoring soil pH over time, this means that one should preferably sample in the same period in the year for the different sampling rounds.

De Gruijter et al. (2006) provide an extensive overview of items that one should consider when designing a sampling scheme. Most importantly is a detailed analysis and specification of the objective. This includes clear definitions of:

- *target universe:* This is a precise definition of the universe of interest (i.e., the *population*: the area that one wants to sample) with boundaries in space and/or time and possibly a specification of exclusions. For instance, the topsoil of all arable fields in a region, the soil up to 1 m depth for areas under forest in a country.
- *domains of interest:* a specification of the part(s) of the universe for which separate results are required for reporting. This could be the entire region, different land cover classes that are found within the target universe, administrative sub-divisions within a country, for instance.
- *target variable*: variable(s) to be determined for each of the sampling units (e.g., soil properties or soil quality indicators that can be derived from measured properties).
- *target parameter*: type of statistic for which a result is needed (e.g., mean, total, fraction, median, trend parameter).
- *target quantity*: combination of a domain, target variable and target parameter. For instance, the mean (parameter) soil organic carbon content in the 0-30 cm layer (target variable) of the forest soils in Sweden (domain).
- *type of result*: qualitative (mode of inference is testing, classification or detection; for instance to determine if the target quantity exceeds a threshold) or quantitative (mode of inference is estimation or prediction).

4.2.1 Sampling for estimating spatial means, totals and fractions

To estimate global quantities, such as means, totals and (areal) fractions, it is generally accepted that the design-based approach, involving probability sampling and design-based inference, is more suitable than model-based approach (Brus & de Gruijter, 1997). In probability sampling, the inclusion probabilities of the sampling units are non-zero for all population units (locations), and known (by design. They do not need to be equal as long as they are known. Sampling locations are selected using a random number generator. The inclusion probabilities are the basis for statistical inference. There are many different types of design-based methods that are all suited to estimate global quantities for a target universe and its domains. Each design-type has its advantages and disadvantages and choosing a design involves many aspects. De Gruijter et al. (2006, p78-79) provide a decision tree to aid one in the choice of a design type for global quantities in space.

The design types described here are more extensively described by De Gruijter et al. (2006), from which most of the text is borrowed, including mathematical notations for the statistical inference. Practical applications with code examples can be found in Brus (*in prep*) or Brus (2019).

4.2.1.1 Simple random sampling

The most basic design type is simple random sampling. This design can be used when absolutely no knowledge of any trends, patterns or correlations in the target area and variable is available. The main drawback of this method is its inefficiency.

With simple random sampling, all sampling locations are selected with equal probability and independently from each other. This sampling is designed to enable simple and straightforward statistical analyses. However, because of the often uneven spatial coverage of sampling locations of this type of design, and because no prior information on the spatial variation is used, the sampling variance is relatively large. Many sampling locations are needed for accurate estimation of global quantities. Moreover, the suitability of this design for estimating quantities of subareas (domains within the target universe) is restricted because one does cannot control selection of sampling sites. It could thus happen that there are domains without any sampling locations, especially when the target universe is large and the total sample size limited. The design is therefore not very well suited if one requires domain estimates besides quantities for the entire target universe. However, the design is very flexible in the sense that the sample size can be easily adapted.

4.2.1.2 Stratified simple random sampling

There are two reasons to divide a target area into sub-areas, i.e. to make use of stratification: i) to improve the accuracy of estimated means and totals for the area as a whole, and ii) to obtain separate estimates for domains of interest that coincide with the strata. With stratified simple random sampling the target area is divided into sub-areas, called strata. Within every sub-area, or stratum, simple random sampling is then applied. This leads to a smaller sampling variance of the estimated global quantities at the same number of sampling points (sample size) is in simple random sampling, or smaller sample size for the same sampling variance of the estimated mean. Stratified random sampling is thus more efficient than simple random sampling. This is achieved by forming strata that are all as homogeneous as possible regarding the soil property in which we are interested. Stratification offers the possibility to control the sampling density within strata based on expected variability, and allows to estimate the global quantities of interest for each stratum as well should the strata be of interest as reporting units. The latter cannot be guaranteed by simple random sampling since one cannot control sample size for subareas within the target area. Another advantage of stratified simple random sampling is that one can take ‘accessibility costs’ into account by stratifying according to accessibility.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

Sampling densities can be lower for strata that are more difficult to reach compared to strata that do not suffer from accessibility constraints. This might result in an 'uneven' allocation of sampling units across strata (i.e., allocation is not proportionate to stratum size) but this is automatically taken care of in the statistical inference so that unbiased and valid estimates of global quantities and their confidence intervals are still obtained. See Kempen et al. (2009) for a practical application of stratified simple random sampling for validation of a digital soil map.

How the target area is subdivided into strata determines for a large part the efficiency of the stratified sampling design. The goal of stratification is to minimize target variable variance within strata and maximize the difference in means among the strata. Two different stratification methods are available to choose from. Stratification by ancillary variables and compact geographical stratification.

Stratification by ancillary variables

If prior knowledge is available as a map of an ancillary variable that is related to the variable(s) of interest (for example a digital elevation model or satellite imagery), this variable can be used to construct the strata. Using ancillary variables, two methods of stratification are possible. Strata can be based on an a priori classification that is based on knowledge about the correlation between the ancillary variable and the target variable(s). Think, for instance, of a soil class map, the units of which are used as strata to estimate the mean of soil properties (Brus, 1994). Alternatively, when dealing with quantitative ancillary variables like remote sensing data, strata can be calculated with, for instance, the cumulative-root-frequency method in case we have a single quantitative stratification variable (Dalenius and Hodges (1959) or cluster analysis like k-means clustering (Hartigan, 1975) if we have multiple stratification variables.

With both qualitative and quantitative ancillary information it is relevant to consider the possible relation they might have to the target property or target properties, as many soil survey will aim to measure multiple soil properties (which can have different spatial variation at the same time. For soil sampling the relation between ancillary information and the target (soil property) variables are often based on the model of soil forming factors, first described by Jenny (1941) and later elaborated by McBratney et al. (2003) in the SCORPAN model. Both are elaborated in more detail in chapter 5 in the sections on covariates for mapping.

Compact geographical stratification

When ancillary variables are not available, the stratification can be based on spatial coordinates through the definition of compact geographical strata that cover the sampling area. Sampling locations are then selected in each of the strata ensuring that these are well spread across the sampling area. The accuracy of the estimated global quantities will usually be increased by dispersing the sample locations so that they cover the study area as uniformly as possible (Brus et al., 1999). For compact geographical stratification to work properly, it is important to ensure that the area is split into sub-areas that are geographically as compact as possible. This can be achieved using a k-means algorithm on spatial coordinates. For design-based estimation it can be convenient to construct compact geographical strata of equal size which is supported by R package spcosa (Walvoort et al., 2010).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

4.2.1.3 Systematic random sampling

A simple way of drawing random samples with good spatial coverage, i.e., samples whose locations are spread uniformly over the study area, is random grid sampling. The grid-spacing is chosen such that the expected number of sampling locations is affordable, or the accuracy of the result is sufficient. In regular grid sampling the pattern of observations is fixed (square, triangular). In general, systematic random sampling gives more accurate estimates of the mean than simple random sampling but no unbiased estimator of the sampling variance is available (i.e., the precision of the estimated global quantities). Furthermore, the target area is represented in a homogeneous way. Brus and Saby (2016) tested different approximation approaches. An operational disadvantage might be that the total travel distance between sampling locations is relatively long because of their even spreading.

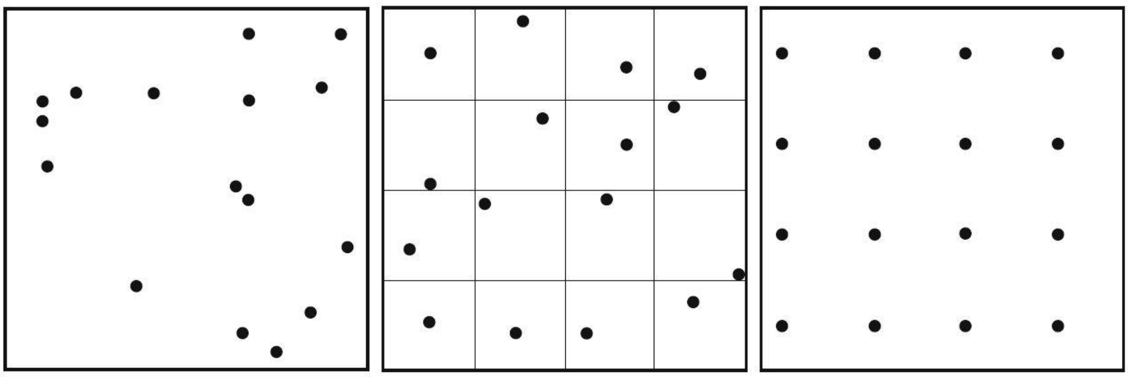


Figure 4.1 National examples of sampling patterns based on simple random sampling (left), stratified random sampling (middle), systematic random sampling (right). Adapted From De Gruijter et al. (2006).

4.2.1.4 Cluster random sampling

In stratified simple random sampling with compact geographical strata and in systematic random sampling the sampling units are well spread in geographical space. In general this leads to an increase of precision of the estimates. With large study areas the price to be paid for this is long travel times, so that fewer sampling units can be observed in a given survey period. In this situation it can be more efficient to select spatial clusters of sampling units. The clusters are not subsampled, in contrast to two-stage random sampling (see below). For this reason cluster random sampling is also referred to as single-stage cluster random sampling. In soil survey, a popular cluster shape is a transect. The reason for using transects is that the individual sampling units of a transect can easily be located in the field, which was a big advantage in the pre-GPS era. The selection of sampling locations for cluster random sampling is not straightforward. De Gruijter et al. (2006, p. 100) describe a selection technique. See Stoorvogel et al. (2009) for a practical application.

4.2.1.5 Two-stage random sampling

In two-stage random sampling, sampling locations are selected in two stages. The area is divided into a number of sub-areas as with stratified simple random sampling. Whereas in stratified simple random sampling all strata are sampled, in two-stage random sampling a limited number of sub-areas is sampled. In the first stage, a limited number of sub-areas is selected randomly. These sub-areas are called the *primary units*. These primary units are preferably selected with probabilities proportional to their size (area of pu's). In the second stage, a set of sampling locations is randomly selected, for instance by simple random sampling, within each of the primary sampling units. The difference with (single-stage) cluster random sampling of the previous section is that here not all sampling units of the selected clusters are observed, but only some of them. For a more detailed description of two-staged random sampling, we refer to De Gruijter et al. (2006, Subsection 7.2.5). A two-stage design can be extended to more stages following the same selection principles. A practical example of a three-stage random sampling is the sampling scheme that is being designed at the time of writing (March 2021) for the EU H2020 Soils4Africa project (paragraph 4.1.4). A report describing this design is expected in the course of 2021 and, once published, can be found on the project's website.

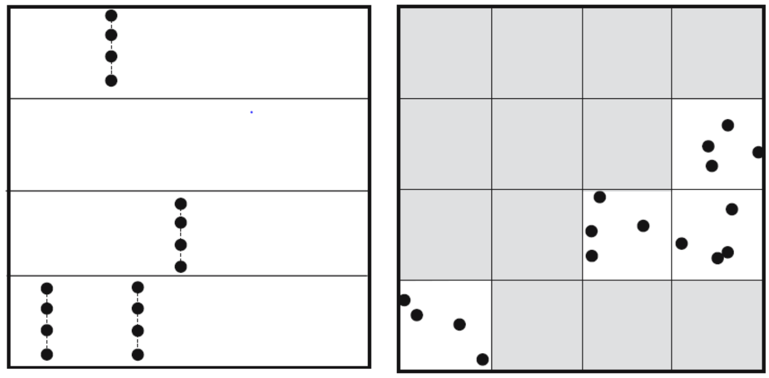


Figure 4.2 Example of a cluster random sample (left) and two-stage random sampling (right). Adapted from De Gruijter et al. (2006).

4.2.1.6 Other probability sampling designs

There are many more probability sampling methods that are designed for very specific purposes. These are able to answer specific questions, but often come more complex statistical inference. However, sometimes the circumstances force the application of these types of designs. The sampling strategies described in the previous subsections can be combined to compound strategies, for example two-stage random sampling with systematic random sampling in both stages, and stratified cluster random sampling. For a description of compound strategies and advanced strategies we refer to De Gruijter et al. (2006, p. 106-110). If information on an ancillary variable is exhaustively available, then gain in accuracy can be achieved by selecting points with probabilities-proportional to this ancillary variable, provided that the ancillary variable is proportional to the variable of interest (Brus et al., 2006b; De Gruijter et al., 2006, Subsection 7.2.9).

For a description of other advanced designs such as sequential random sampling, adaptive cluster sampling, two-phase random sampling we refer to De Gruijter et al. (2006, Subsects. 7.2.10-14).

Balanced sampling

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

In balanced sampling a linear relation between a target variable and one or more covariates is exploited in selecting the sampling locations. It is efficient to select a sample so that the sample average of the covariate equals the mean of the covariate in the population. We refer for a more detailed discussion on balanced sampling to Brus (2015).

4.2.1.7 Model-assisted methods for estimating means and totals

An alternative approach for estimating means and totals for the entire area or several subareas is the model-assisted approach. In the model-assisted approach the sampling locations are selected by probability sampling, and the inclusion probabilities are used to estimate the mean or total. This is the same as in the design-based approach. However, the mean or total is not estimated by the usual Horvitz-Thompson estimator, but by a more efficient estimator that exploits the relation of the study variable with one or several covariates. Well-known examples are the regression estimator, ratio estimator and the post-stratified estimator. The role of the model is entirely different from the role in the model-dependent approach (see hereafter) (Brus 2021). In the model-assisted approach we do not rely on the modelling assumptions, the model is only used to derive an efficient estimator. Even if the model is not correct, model-assisted inference yields (approximately) design-unbiased estimates of means and totals and correct coverage rates of confidence intervals (Brus, 2000). The usual regression estimator requires that the population mean of the covariates must be known. If these are unknown, a two-phase sampling approach can be used, see Brus and Te Riele (2001) for an application in mapping groundwater table depths.

4.2.1.8 Model-based prediction of means and totals

Even though probability sampling is in many ways favorable so that global quantities can be estimated by design-based or model-assisted inference, in case we have a non-probability sample (inclusion probabilities are unknown) global quantities can still be predicted by model-based inference. The covariance structure of the error is estimated and then used to estimate the quantities with known variance. Residual maximum likelihood (REML) remains the best way to estimate the variance parameters since it is unbiased. Lark and Cullis (2004) describe in detail the approach for a systematic sampling grid. Note that if locations are selected by probability sampling global and local quantities can be estimated by design-based inference, model-assisted inference, or model-based inference. In this sense probability sampling is most flexible.

4.2.2 Sampling for mapping

When not global quantities, but local quantities (i.e., predictions at points) in space are needed, other methods are available since design-based approaches involving probability sampling are less attractive for this purpose. Therefore designs such as geometric, adapted experimental or model-based sampling designs are needed to make local estimates or predictions, for instance using digital soil mapping (DSM), see Brus (2019) for a detailed discussion and for R scripts supporting these designs. When choosing a design it is important to already know which mapping method will be used later on, once the data is collected. The 'optimal' sampling design for a given survey strongly depends on the aim but also on the model used. For example, for ordinary kriging we will use a very different sampling design than for regression kriging or machine learning. An overview of Digital Soil Mapping and Conventional Soil Mapping methods principles and its pro's and con's are detailed in chapter 5.

4.2.2.1 Geometric sampling designs

In geometric sampling designs a criterion is minimized that is a function of the distances of the sampling points to the nodes of a fine discretisation grid. These distances can be measured in either geographic or feature (attribute) space. Feature space or attribute space refers to the n-dimensions where your variables live, for one variable this will for instance be the space between its minimum and maximum. Examples are spatial coverage sampling (Walvoort et al., 2010) and feature space coverage sampling using hard k-means (Wadoux et al., 2019). In adapted experimental designs, samples are selected to cover the multivariate feature space that will be used for DSM, as in a factorial field experiment. Examples are conditioned Latin hypercube sampling (Minasny and McBratney, 2006) and response surface sampling (Lesch, et al., 1995).

Centred grid sampling

A straightforward, popular sampling method for mapping is sampling on a regular grid, for instance a square or triangular grid. As opposed to regular grid sampling in a design-based approach (systematic random sampling), in a model-based approach there is no need to place the grid randomly on the area, but can be placed in such way that it is most convenient or has optimal coverage of the study area.

When sampling on a regular grid we must decide on the grid-spacing, i.e. the distance between neighbouring points. This boils down to a decision on the sample size, i.e. the number of grid points. There are two options to decide on this spacing, either by starting from the available budget or from a requirement on the quality of the map. In principle, the higher the map quality the larger the number of observations needed. In some methods, such as the model-based methods (section 4.2.2.3) the relation between map quality (expressed in the form of the interpolation error variance) and sampling density can be quantified which allows one to derive the needed sampling density for a required map quality.

Spatial coverage sampling

With regular grid sampling of irregularly shaped areas the geographical spreading of the sampling locations throughout the study area can be suboptimal. In this case, we would like to relax the constraint of sampling on a regular grid. We would like to shift grid points a bit into the under-sampled areas (for instance occurring at the boundary of a survey area), so that the spatial pattern becomes irregular. This leads to spatial coverage sampling in which a geometric criterion defined in terms of the distances between all locations in the survey area and the sampling points is minimized. Such a sampling design can be considered optimized in geographical space (Brus et al., 1999; Brus et al., 2006a).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Feature space coverage sampling

In centred grid sampling and spatial coverage sampling no covariates are used in designing a sample. If (gridded) maps of covariates are available that are most likely correlated with the variable of interest (chapter 5) it is attractive to use these maps in designing a sample for mapping. This can include soil maps, elevation maps, satellite imagery or proximal sensing data when available for the entire area. One option is the same as we used before to construct strata for design-based estimation of means, totals and fractions. The raster cells of the maps are clustered into as many clusters as we want to select sampling points. Once the clusters are constructed, from each cluster one sampling point is selected, not randomly as in probability sampling for design-based inference, but purposively, closest (in feature space) to the centroid of the cluster.

Clustering is done in such a way that the average distance of the population units (raster cells) to the nearest sampling unit is as small as possible in the feature space as defined by the prescribed set of environmental covariates. By taking the squares of the shortest distance, the criterion can be minimized by the *k*-means clustering algorithm. In feature space coverage sampling, the feature values need to be standardized due to the fact that the ranges of these features are largely different from each other. An example is for instance the values for elevation and slope. The standardization is done by first subtracting the mean and then dividing by the standard deviation.

4.2.2.2 Adapted experimental designs

Response surface designs

In response surface sampling using a central composite design the locations are optimized for calibrating a second-order polynomial regression model for the variable of interest, using multiple covariates as predictors. Predictors are covariates that are expected to have a relation with the variable of interest, for instance because to relate to soil forming factors (chapter 5). This can include soil maps, elevation maps, satellite imagery or proximal sensing data when available for the entire area. It is a method often used to define sampling locations needed to calibrate a model to estimate soil properties from either remote or proximal sensing data, such as satellite imagery, UAV imagery or proximal measurements using EMI or gamma radiation (chapter 5). In a regression model it is assumed that the data are independent, and for that reason the locations are spread in geographical space. Further, in experimental design, the factors are independent, whereas in observational studies the covariates can be correlated. For that reason principal components of the covariates are used in the sampling design.

Conditioned Latin hypercube sampling

Conditioned Latin hypercube sampling (CLHS): CLHS selects samples in such a way that the marginal frequency distributions of the quantitative features (or environmental features) in the population are reproduced by the sample as closely as possible (Minasny and McBratney, 2006). To this end for each feature, marginal strata are constructed using equally spaced quantiles of the marginal distribution, so that all marginal strata contain an equal number of pixels. Recently Wadoux et al (2019) and Ma et al. (2020) compared feature space coverage sampling and conditioned Latin hypercube sampling for mapping quantitative soil attributes and soil classes, using a random forest for mapping. Surprisingly conditioned Latin hypercube sampling did not perform well, and feature space coverage sampling outperformed cLHS.

4.2.2.3 Model-based sampling

Model-based sampling approaches are of interest when one aims to use a geostatistical method for mapping such as the various forms of kriging. In model-based sampling, a geostatistical model is used at the stage of designing a sample. So contrary to the geometric designs and adapted experimental designs this approach requires prior knowledge of the spatial variation (variogram). The model is used to optimize the sampling locations. Once the data are collected, these data are used to fit a geostatistical model which is subsequently used for mapping. The model used for mapping will not be the same as used for designing the sample.

If the variogram[175] of the property of interest is known then it is possible to optimize the sampling scheme such that an objective function related to the spatial interpolation error is minimized. Average kriging variance or maximum kriging variance are most often used as such criteria to be minimized. A tool called Spatial simulated annealing (SSA) for optimal spreading of the sampling locations either in geographical space (simple or ordinary kriging variance) or geographical space as well as in feature (or attribute) space (universal kriging variance) was introduced by Van Groenigen and Stein (1998). In short, the SSA algorithm is a combinatorial optimization algorithm that optimizes a sampling configuration iteratively, by shifting points in arbitrary directions by a defined distance, which becomes shorter as the algorithm proceeds. The optimization criterion is the interpolation error variance. If the new configuration shows an improvement (lower variance) the particular point is kept in its new position, if not it returns to its former position. Ending at a local minimum is prevented for by incorporating a random element. In practise, thousands to hundreds of thousands of iterations are performed.

Although the variogram is rarely known prior to sampling, it can be estimated by expert elicitation (Truong et al., 2013). Moreover, SSA can be extended to optimize the sampling design for multivariate soil mapping (Vašát et al., 2010). On the other hand, Wadoux et al. (2019) showed that the direct optimization of spatial sampling designs was only rarely worthwhile. For most cases, it is best to apply a spatial coverage scheme with a proportion of additional sampling locations to provide some closely spaced pairs if a variogram must be estimated. One advantage of spatial coverage schemes is that they do not depend on the variogram of the soil property to be sampled. Coverage schemes are created by minimizing a criterion that is simply a function of the distance between sampling locations (see 4.2.2.1). A practical solution, as suggested for instance by De Gruijter et al. (2006, pp. 166-168), is to supplement the spatial coverage sample by a few additional units (about 50), located at short distances from the existing units should a variogram need to be estimated.

Model-based sampling can also be extended to the situation where one or more maps are available of covariates related to the variable of interest. Using a geostatistical model in which the mean is

[175] See Webster and Oliver (2007) for more an explanation about variogram modelling and kriging.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

modelled as a linear combination of covariates for designing a sample, results in samples that are automatically well-spread both in geographical space and in feature space (Brus and Heuvelink, 2007).

4.2.3 Sampling for monitoring

This section addresses sampling for monitoring in both space and time. One of the main purposes of monitoring is to estimate the *change* of a (soil) variable between two sampling times (cycles). For instance, we may be interested in the change in the mean soil organic carbon concentration in the topsoil of a country, or the change in areal fraction of degraded soil in a landscape after intervention measures have been put in place. If one has more than two sampling times, then the interest might be in the average change per time unit of the mean, total or fraction, i.e., the temporal trend. Typically estimated global target quantities in space-time are (De Gruijter et al., 2006):

- the current mean (i.e., the spatial mean at the most recent sampling time);
- the change of the spatial mean from one sampling time to the other;
- the temporal trend of the spatial mean;
- the spatial mean of the temporal trend;
- the spatio-temporal (ST) mean;
- the difference between the ST means before and after an intervention.

In this report we only consider sampling methods for monitoring global quantities (means, totals, fractions) in space-time. We believe these are most relevant for within the context of EJP Soil and the design of soil monitoring networks. Methods for quantifying local (i.e., location-specific) quantities in space-time are not considered here. These methods pertain to spatio-temporal mapping, i.e., mapping a target (soil) variable over the entire space-time universe and updating maps. We refer the reader to De Gruijter et al. (2006) who provide an elaborate overview of (model-based) methods for spatio-temporal mapping such as space-time kriging and Kalman filtering. See for instance Heuvelink and Griffith (2010) for an application of space-time geostatistics and Heuvelink et al. (2006) and Webster and Heuvelink (2006) for applications of Kalman filtering. Heuvelink et al. (2013) provide an example of sample design optimization for space-time kriging. More recently, Aktar et al. (2021) provided an application of space-time modelling with machine learning. More general reviews of soil monitoring are provided by Arrouays et al. (2012) on generic issues on broad-scale soil monitoring networks and Morvan et al. (2008) on the different soil monitoring networks that exist in Europe and how these could be harmonized (also see Section 4.3.2).

4.2.3.1 Statistical sampling approaches for soil monitoring

In designing a sample for monitoring, not only spatial variation is relevant, but temporal variation must be taken into account as well. This means that sampling times must be selected in addition to sampling locations, leading to four possible sampling approaches, i.e., combinations of probability (P) and non-probability sampling (NP) (Brus, 2014):

- P-sampling in both space and time (P + P),
- NP-sampling in both space and time (NP + NP),
- P-sampling in space and NP-sampling in time (P + NP), and
- NP-sampling in space and P-sampling in time (NP + P).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

With P + P the statistical parameter of interest (mean, total, fraction) can be estimated with a fully design-based method (Brus, 2014). This means that a model of variation in space and time is not needed for statistical inference that is fully based on the spatial and temporal sampling designs (Brus and De Gruijter, 2012). The P + P approach is advantageous in compliance monitoring of the space-time mean or space-time total (Brus and De Gruijter, 2012), such as for instance the total annual CO2 emission in a country or the change in total SOC stock in a region. Brus and Knotters (2008) provide an application of P + P for compliance monitoring of water quality and Brus et al. (2010) for estimating average annual fluxes of chemical elements from forest soils. The fully design-based approach avoids using models (and hence model assumptions) which enhances the validity of the result, which is important for compliance monitoring.

In the NP + NP approach a stochastic (statistical) model of the variation in space and time must be postulated to estimate the parameter of interest (i.e., mean, total, fraction) (Brus, 2014). This means that the statistical inference is fully model-based. De Gruijter et al. (2006) explain that fully design-based methods are generally well suited for estimating these properties, but that in some cases using fully model-based methods could be advantageous. For instance, this might be the case in a situation where one has prior monitoring data from a purposive (i.e., non-probability) sample in space-time that needs to be extended with additional data. Since NP + NP requires models of variation in space-time, these methods are more sophisticated than methods for a fully design-based approach, and require a solid understanding of geostatistics. De Gruijter et al. (2006, Section 15.3) provide an extensive overview of geostatistical methods for a fully model-based sampling approach.

The P + NP approach uses a hybrid estimator. In this hybrid approach sampling locations are selected by probability sampling, whereas sampling times are not. Brus and De Gruijter (2012) provide a theoretical overview of the P + NP approach with a focus on the estimation of the temporal trend of the spatial mean that they illustrate with a case study on forest soil eutrophication and acidification. For this case study sampling locations were selected with probability sampling, sampling times were selected preferentially and a model of the temporal variation of the spatial means was postulated (a time series model). Brus and De Gruijter (2012) show how to do the inference of the P + NP approach with sampling repeated at constant time intervals.

According to Brus (2014) the NP + P approach with non-probability sampling in space and probability sampling in time is rarely used in practice and is therefore not considered here.

Within the context of EJP SOIL and designing soil monitoring networks with the aim to provide estimates of global quantities in space-time, especially the P + NP and the P + P are most relevant approaches to consider. In both approaches sampling locations are selected by probability sampling. We note here though, that probability sampling in time is less common than probability sampling in space (De Gruijter et al. 2006). This is not so strange for practical reasons. Fieldwork costs may increase a lot if we need to go into the field at many different points in time to collect perhaps not that many soil samples

4.2.3.2 Types of sampling patterns for monitoring

The efficiency of the monitoring design is both determined by the distribution of the sampling locations and by the distribution of sampling moments of these locations. Monitoring methods are designed to minimize this combined variation and the costs of sampling. Given any of the basic spatial designs discussed in Sections 4.2.1 and 4.2.2, two different kinds of temporal restrictions can be imposed to these designs to increase the efficiency for monitoring purposes. Either sampling locations are visited multiple times (stationarity), or multiple sampling locations are visited simultaneously (synchronicity). By imposing these restrictions in different combinations, four different sampling patterns arise.

Static sampling revolves around the principle of static sampling locations. This means that sampling takes place at a fixed set of locations that are revisited (Figure 4.3a). Sampling at the various locations may or may not follow the same pattern in time (De Gruijter et al., 2006). *Synchronous sampling* designs, also known as repeated or dynamic sampling, revolve around the principle that a different set of sampling locations is selected for each sampling time (De Gruijter et al., 2006) (Figure 4.3b). Note that synchronous sampling is sometimes referred to as *independent synchronous sampling* (e.g. Brus, 2014).

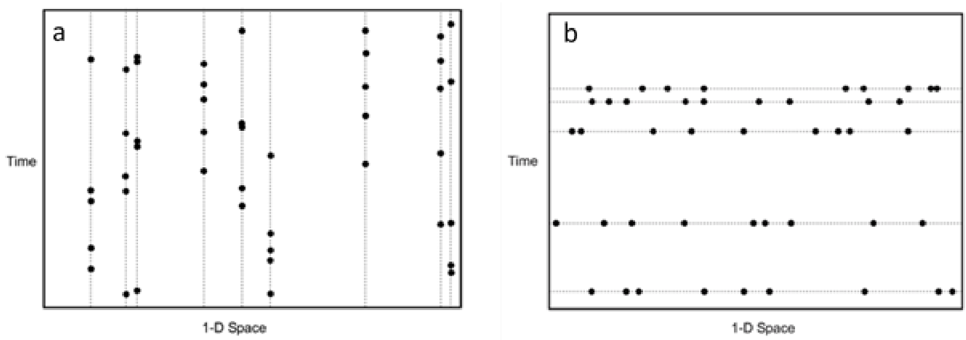


Figure 4.3 Schematic visualisation of a static design with simple random sampling in both space and time (a) and a synchronous design with simple random sampling in both space and time (b). (Reproduced from De Gruijter et al., 2006).

Combining a spatial sampling design and a temporal sampling design results in a *static-synchronous sampling* design (Figure 4.4a). At every selected sampling moment in time, all sampling locations are visited. Such a design is also referred to as a pure panel. The fourth option is *rotational sampling* (Figure 4.4b), which is meant as a compromise between the rigid, unbalanced static design and the relatively inefficient synchronous design. The rotational design differs from the static-synchronous design in that the locations of the previous sampling time are partially replaced by new ones (De Gruijter et al., 2006). Sometimes this is referred to as 'sampling with partial replacement'.

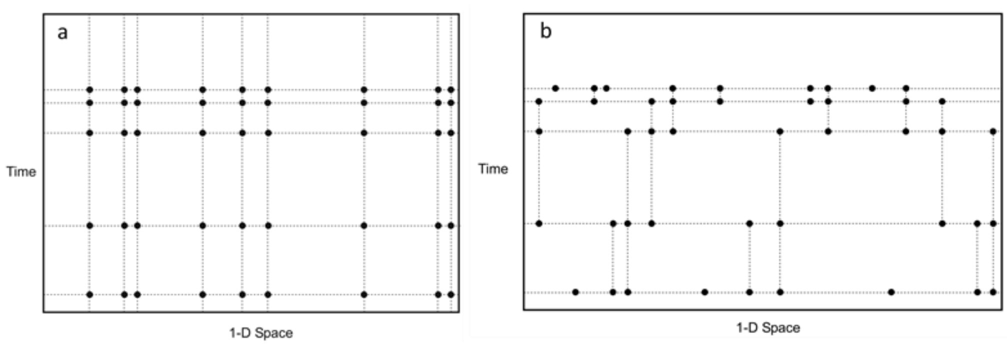


Figure 4.4 Schematic visualisation of a static-synchronous design with simple random sampling in both space and time (a) and a rotational design with simple random sampling in both space and time (b). (Reproduced from De Gruijter et al., 2006).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

In addition to these four basic sampling pattern types, two variations on these designs are worth mentioning there. The first is the *serially alternating design* (Brus and De Gruijter, 2011; Brus and De Gruijter, 2013). In serially alternated sampling, disjoint (i.e., non-overlapping) sets of sampling locations are selected and then observed in turn in a cyclic fashion (Brus and De Gruijter, 2011). The second is the *supplemented panel design*. This design is a compromise between a synchronous and static synchronous design. One set of locations (the 'pure panel' part) is maintained through time, supplemented with a different set of locations each time (Brus and De Gruijter, 2011; Brus and De Gruijter, 2013). Figure 4.5 provides notional examples of five space-time sampling designs that show how the different panels (different sets of sampling locations) are used in each design. In section 4.2.3.4 some guidance is given for choosing a space-time design.

4.2.3.3 Sampling strategies for monitoring

Once the space-time sampling pattern has been chosen, one must choose a strategy for selecting the sampling locations as well as the sampling times. As we have seen in Section 4.2.3.1, probability sampling as well as non-probability sampling can be used for this purpose. Here we advise to select sampling locations by a probability sampling design. A brief overview of these designs was given in Section 4.2.1 and a more elaborate description is given in Sections 7.2 and 15.2 of De Gruijter et al. (2006). Probability sampling in time is less common than in space (De Gruijter et al. 2006). Nevertheless, most design-based methods presented in Section 4.2.3.1 and Section 7.2 in De Gruijter et al. (2006) can be used for sampling in time. Typically, sampling times are chosen purposively. This results in a hybrid (P + NP) sampling approach. This does, however, not mean that the statistical inference is done with a hybrid estimator. In many cases, also for soil monitoring, design-based estimators can be used in a P + NP sampling approach. This is for instance the case when the target quantities such as the change of the mean (total, fraction) between two time periods, the space-time mean and the temporal trend of the spatial mean that can be defined in such a way that design based inference is possible (see for instance Brus and De Gruijter (2011) and Brus (2013)). In that case the target universe is restricted to the sampling times only; what happens between the sampling times is disregarded.

The simplest and most common model-based method of sampling in time is systematic sampling; this means sampling at constant time-intervals (De Gruijter et al., 2006). A time-series model can then be used for statistical inference from the sample data (see Box and Jenkins, 1976 for examples and Appendix C in De Gruijter et al., 2006). Systematic sampling requires to choose an appropriate sampling frequency. Sampling at constant time intervals has operational advantages and the methods used to analyze are mathematically relatively simple (De Gruijter et al., 2006). However, De Gruijter et al. (2006) warn that systematic sampling in time is not always the best option. For instance, if temporal variation varies with time then it is more efficient to sample more densely in periods with larger variation, i.e., one might then want to stratify over time. Another problem might be if there is temporal periodicity in time and sampling frequency coincides with the period. For instance, to estimate the monthly topsoil temperature it would be unwise to measure once a day, always at the same time.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

4.2.3.4 Designing a monitoring scheme: flexibility, robustness and operational aspects

De Gruijter et al. (2006) stress an important difference between sampling for estimating global quantities and mapping on the one hand and sampling for monitoring on the other hand. The former typically takes place in a relatively short period of time during which the universe of interest as well as the operational and budgetary aspects do not change. With monitoring not only the universe may undergo (large, unexpected) changes but also the sampling conditions, especially in long-term monitoring. This makes that adaptation of a (long-term) monitoring scheme are inevitable or at least desirable. Budgets may vary from year to year and operational constraints that were initially present might disappear or vice versa. Other changes that might occur over time that affect the monitoring scheme are the definition of new objectives (e.g. target variables, domains of interest) or better measurement techniques become available. Finally, more and more data become available about the universe of interest that can give good reason to fine-tune or redesign the scheme (De Gruijter et al. 2006). To be able to adapt to (unforeseen) changes in conditions calls for flexibility of the monitoring scheme and the original design should be such that it allows flexibility.

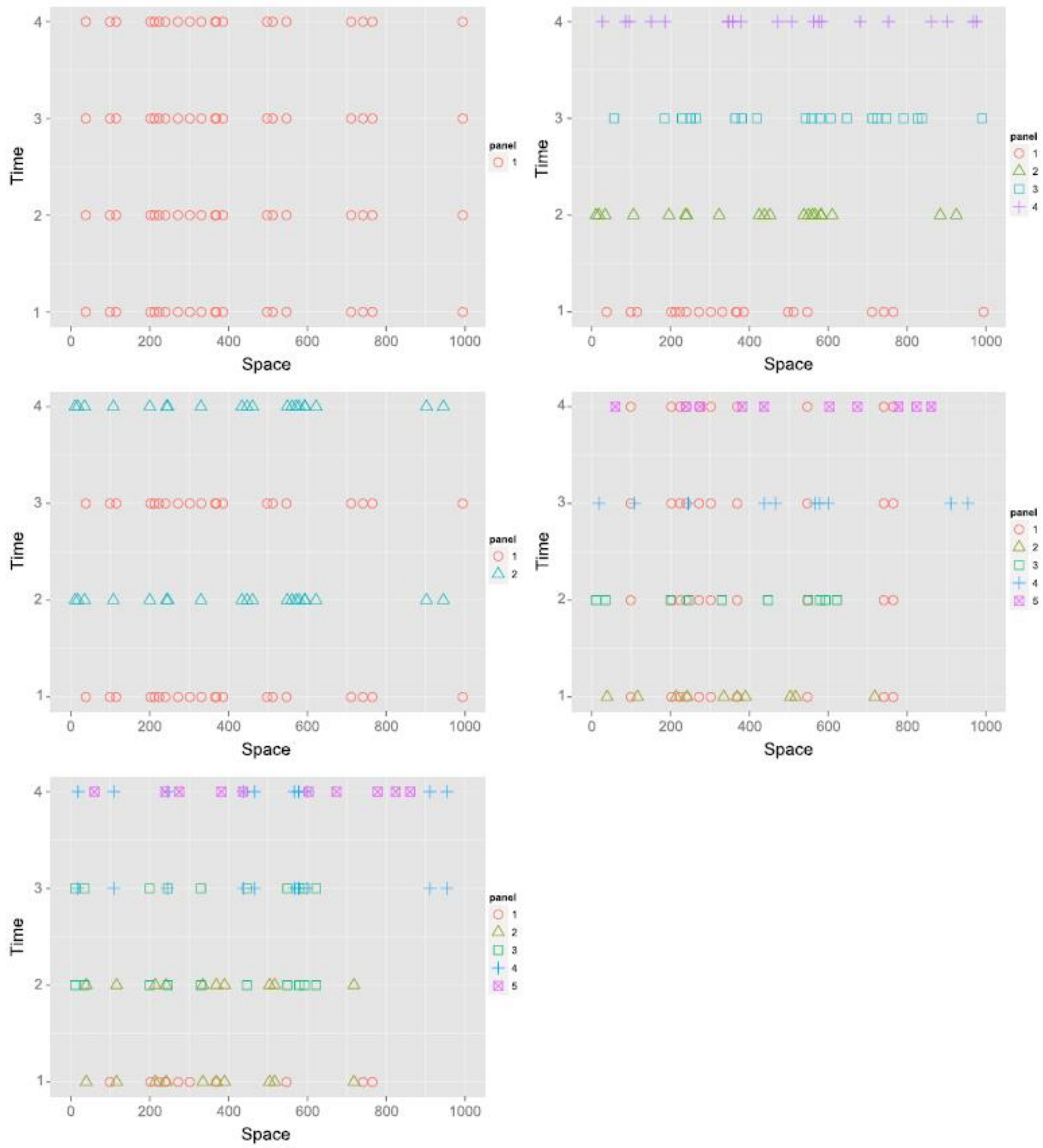


Figure 4.5 Notional examples of five space-time designs. Static synchronous (top left), (independent) synchronous (top right), serially alternating (middle left), supplemented panel (middle right), rotational (bottom) (Reproduced from Brus and De Gruijter, 2011).

Overton and Stehman (1995) discuss design implications and how the choice of a design is influenced by intended users of the data, whereas Overton and Stehman (1996) elaborately discuss desirable design characteristics for long-term monitoring of ecological variables and ways to adapt the design to changing conditions, including sample restructuring, changing the size of a sample (increase or reduce) and post-stratification. These authors stress the importance of keeping a design as simple as possible. Overton and Stehman (1995) note that complex sampling designs require complex computational formula for statistical inference. By keeping a design simple, standard statistical analyses commonly used by environmental scientists are correct or at least adequate. Furthermore, they note that the primary concern when designing a monitoring scheme should be to develop an adequate design that

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

makes good use of the available resources and not to construct the perfect, optimal design. Practical convenience and simplicity cannot be sacrificed to achieve optimal statistical efficiency (Overton and Stehman, 1995). To avoid (overly) complex and less adaptive designs Overton and Stehman (1996) suggest to use sample structures (such as stratification) only sparingly and use equal probability designs. Examples of these are simple random and stratified sampling with proportional allocation.

It is a misconception that for monitoring global quantities (means, totals, fractions) in space-time one should always return to the same sampling location during each sampling cycle. For estimating the change of the mean, total or areal fraction from one sampling time to another revisiting all sampling locations is the best option. However, as will be explained hereafter, for other space-time quantities revisiting it can be more efficient to revisit a subset of the sampling locations only. As we have seen in the previous sections and Figure 4.5, different sets of sampling locations can be used in different sampling rounds such as in the synchronous design, or combinations of sets of sampling locations such as in the supplemented panel or rotational design. The fact that one could select new sampling locations for each sampling cycle gives flexibility to the scheme and reduces the chance that a sample location cannot be visited anymore in a next sampling round (moreover we will rarely be able to go back to the exact location), for instance when the site is used for construction or when access permission by the land owner is denied. However, if one does revisit (a selection of) sampling sites in subsequent sampling rounds, as will be the case in most sampling designs, seeFigure 4.5, and a site cannot be accessed or sampled anymore, then this does not mean the statistical integrity of the monitoring scheme is compromised. Statistical inference still yields valid and unbiased estimates of global quantities.

Besides `ad-hoc` changes in sample size as a result of accessibility issues, a more structural reduction in sample size can be required, for instance after a budget cut. Reducing the number of sampling locations to be revisited is straightforward when the existing sample is selected with probability sampling. One can simply take a probability sample of the original sample since a probability sample of a probability sample is itself a probability sample of the sampling universe (Overton and Stehman, 1996). Design-based estimates are still possible with the only difference that the results will be less precise (have larger variance). Like reduction, increasing the size of a probability sample is in most cases straightforward (Overton and Stehman, 1996).

When designing a scheme for long-term monitoring one must anticipate that adaptions will be required over time. Some smart design choices at the beginning can increase the robustness of a design and greatly ease adaptions later on. Such choices are an equal probability design, and a design type that allows straightforward adjustment of sample size. For instance, changing sample size for a systematic random sample is less straightforward than that of a simple, stratified or two-stage random sample (Overton and Stehman, 1996). Another simple measure one can take in case of a stratified design is to ensure that a large enough number of sampling locations is selected for each stratum so that a possible reduction of sampling size in the future is less likely to compromise the sample structure. The minimum sample size for a stratum is two (to be able to estimate the stratum variance). By choosing a sample size per stratum large enough at the beginning, one can ensure that after a (substantial) budget cut at least two (permanent) sample sites remain in each stratum. When designing a sampling scheme this could mean that fewer strata must be defined than initially intended to allow allocation of sufficient sampling locations per stratum in order to anticipate possible budget cuts in future. Fewer, larger strata may yield in less precise results at the original budget but the expected loss

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

of initial precision may be less important than greater adaptability to changing budgets (De Gruijter et al., 2006).

Next, we consider some statistical and operational aspects of choosing a space-time design for monitoring as summarized from De Gruijter et al. (2006, Sections 14.2 and 15.2):

- *Static* and *static-synchronous* patterns are attractive when the costs of repeated sampling at the same location are lower than for sampling at different locations with the same total sample size. This is for instance when semi-permanent measuring equipment is installed at fixed locations. A statistical disadvantage of these two patterns is that only information on temporal variation increases and not that on spatial variation, in contrast to synchronous and rotational patterns (incl. serially alternating and supplemental panel). An advantage of static over static-synchronous is that sampling times may be adapted to local circumstances (e.g. sampling at a time when operational costs are lowest), while static-synchronous patterns reduce the number of sampling times given the sample size compared to static patterns. This means that if costs for an additional sampling time are larger than for sampling additional locations at a given sample time, it can be attractive to reduce the number of sampling times and increase the number of sampling locations per sampling time. This enables more locations to be sampled in total, yielding more accurate estimates of spatio-temporal global quantities. A static pattern is attractive when considerable spatial variation between time series is known to exist. The French soil monitoring network is an example of a static-synchronous pattern.
- *Synchronous* patterns are much more flexible than static and static-synchronous patterns. With each sampling time one is free to choose a spatial design (see section 4.2.1 here and 7.2 in De Gruijter et al., 2006) or change the existing design (e.g. sample size, possible stratification, clustering) to adapt to the circumstances at the sampling time. Statistical inference from a synchronous sample is much simpler compared to static, static-synchronous and rotational samples since no sampling locations from earlier sampling times are revisited.
- *Rotational* patterns are more flexible and have better spatial coverage than static and static-synchronous patterns. If there is a fair amount of correlations between observations at consecutive sampling times, then rotational patterns are more efficient in estimating current means, totals and fractions as well as temporal trends than synchronous patterns. A disadvantage of rotational patterns is that design-based inference of the sampling variance is somewhat more complicated.
- *Supplemented panel* designs have the same advantages as a rotational design. Besides supplemented panel has the advantage of operational simplicity: a subset of locations is fixed (the pure panel subset), whereas others are swarming. Besides at the fixed locations a time series of data is obtained.

Finally, we briefly consider suitability of the four main sampling patterns for estimating global quantities (see bulleted list in the introduction paragraph of Section 4.2.3). We limit ourselves here to the 'current global quantities' and 'change of the spatial mean', and the 'temporal trend of the spatial mean'. For a more elaborate overview, including other global quantities, we refer to De Gruijter et al. (2006, Section 15.2). The following summary is adapted from De Gruijter et al. (2006):

- *Current global quantities* (means, totals, fractions): rotational designs are preferred over synchronous or static-synchronous designs. Because sampling locations partially overlap in rotational designs, these designs can exploit information from previous sampling times. Synchronous designs do not have such overlap and there is no simple way to use information from a previous sample to estimate a current quantity. Static-synchronous designs have fully overlapping sample locations (use one panel only) so there is no additional information from in the measurements taken in a previous sampling round.
- *Change of global quantities*. Static-synchronous designs are more efficient than synchronous designs (though the latter are suitable for this purpose as well) because one profits most from the correlation of the two sample sets measured at the two times.
- *Temporal trend of the spatial mean*. With more than two sampling times, a supplemented panel design can be more efficient to estimate the average change of a global quantity per time unit than a static-synchronous design. This depends on how persistent the spatial patterns of the soil variables of interest are (Brus and de Gruijter, 2013). When the spatial patterns do not change much over time (strong persistence) a supplemented panel design well-estimates a linear time trend of the population mean, whereas for highly dynamic variables resulting in large changes in the spatial pattern over time (e.g., locations with high values at a given time may well have small values at a subsequent time) a synchronous or serially alternating design is the best choice. For moderate persistence the choice is more complicated; supplementary panel or serially alternating designs are good choices (Brus and De Gruijter, 2013).

4.2.3.5 Combing different aims

As we have shown in previous sections, the type of sampling design and method of statistical inference depends on the purpose for which the soil data are collected. An important distinction can be made between designs most suitable for estimating global quantities such as means, and areal fractions in space and time and designs suitable for mapping that quantify local quantities (i.e., location-specific predictions) in space. For the former probability sampling designs and design-based inference are preferred, while for the latter probability sampling is not necessary and sampling designs optimize sample locations in geographic or feature space. We realize though that in many cases there is interest in estimating or monitoring global quantities in space(-time) of the target universe or domains of interest within this universe, as well as in using the data for (digital) soil mapping while there are no resources available for two different sampling campaigns.

To serve these different aims we search for a sampling design type that is efficient for both objectives. The efficiency of a sampling design type for DSM largely depends on the mapping method (Brus, 2019). Mapping methods that exploit the availability of maps of covariates related to the soil properties of interest, such as terrain attributes, climate variables and variables derived from remote sensing imagery, are most promising. For these mapping methods spreading of the sampling locations in feature space may increase the efficiency. To ensure that the same data can also be used for design-based estimation of means, totals and areal fractions of (parts of) the universe of interest, we propose

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

to select the sampling locations by probability sampling (see section 4.2.1), using a design type that results in samples that are well-spread in the space spanned by important features. An interesting sampling design type for this is spatially balanced sampling through the local pivotal method (see Grafström et al., 2012; Grafström 2013).

4.3 Additional spatial sampling

There can be various situations in which additional sampling is possible, desirable or needed. If existing data are collected according to a specific sampling design for model-based mapping that serves its purpose and budget becomes available for additional sampling, one can simply optimise the same criterion that one had otherwise chosen but taking as a starting point the data that are already present rather than starting from scratch. In other situations this may not be possible, for instance due to underrepresented areas or feature spaces in a design due to a change in objective, and additional sampling will need to be carried out (4.3.1). Another situation can be that various existing surveys are present that are relevant but which need to be combined for one new objective (4.3.2).

4.3.1 Additional sampling in underrepresented areas: gap filling

For many reasons it may happen that there are gaps (completely unsampled subareas or spots) or sparsely sampled parts of the area of interest in the existing sample design (usually formed by some kind of legacy data). A situation could be that a location cannot be visited because of difficult terrain, locked gate, or other reasons. One can also think of new domains of interest that do not coincide with strata. In these situations the use of legacy data collected for other purpose that could be considered when designing a new survey.

While in the past such a scheme may have served its purpose very well, it may be insufficient for today's applications (e.g. digital soil mapping). Or, with some extra budget, we simply could make an existing sampling scheme denser in order to produce a more accurate soil map. In most of such cases the use of spatial coverage sampling (SCS) (Brus et al., 1999; 2006) —leading to a design that optimizes distribution of samples in geographical space—could be the right choice in sampling for mapping, as this method is capable to take into account existing sampling spots and to spread out the new ones as uniformly as possible to cover the target area evenly. Using the SCS method one can benefit from the delineated geographical strata in several ways. While for monitoring, global quantity estimation purposes or even sampling for validation of digital soil maps (Brus et al., 2011) a simple random sampling within the strata can be appropriate (one random point per stratum), for digital soil mapping application (especially with geostatistical methods such as simple or ordinary kriging) one would prefer to choose the centres of the empty strata as sampling points in order to achieve even geographical distribution.

Moreover, with the information from existing soil samples, whether from previous soil survey or legacy soil data, the use of spatial simulated annealing (SSA) methods seems now more beneficial as the requirement for a known semi-variogram model is far less restrictive in this situation (it can be estimated from existing data) than in the case of a completely unsampled area. The SSA method can be applied for mapping with geostatistical methods such as ordinary kriging. However, one must consider whether the extra portion of work will be worth the map accuracy improvement, which could be only marginal compared to SCS method (Vašát et al., 2010).

So far, we have only considered the case where it was required to spread out the sampling points to cover the geographical space (defined purely by geographical coordinates). However, by applying more advanced soil mapping approaches that benefit from the use of environmental covariates as predictors—for which a relationship to the mapped characteristic is assumed by means of soil forming factors and/or sensing data (chapter 5) — in order to increase the accuracy of final soil maps, one may face the situation that gaps appear in the statistical distribution coverage of the considered ancillary data (in the case of continuous character of the data) and/or in some of the categories coverage (in the case of categorical predictors). In this situation it is required to spread out the sampling points in feature (I.e., attribute space) which leads to a model-based sampling design (paragraph 4.2.2.3). Analogically as in the case of geographical space coverage, here one can use optimization techniques that solve the spatial distribution of soil samples in a feature space such as SSA and/or cLHS and/or feature space sampling. The SSA method is appropriate if the regression (or more generally speaking the universal) kriging is used as a mapping tool. The existing sampling points are fixed at their positions and new points are spread out at the empty or sparsely sampled parts so that criterions such as average kriging variance or maximum kriging variance are minimized. If machine learning is used as a mapping approach, then an even coverage of statistical distribution of predictors is preferred. In that case cLHS (Minasny and McBratney, 2006) and its modifications (Malone et al., 2019; Roudier et al., 2012) can help with achieving the goal as it was designed to coverage the statistical distribution across all considered predictors as evenly as possible for the specified number of soil samples. However Wadoux et al. (2020) demonstrated that feature space coverage sampling outperformed cLHS. For empty or sparsely sampled categories of categorical predictors it would be optimal to have proportionally (with respect to the area of each category) the same number of samples as in other categories. A simple random sampling would be feasible to sample within these empty categories.

4.3.2 Combining two samples

Given the costs of soil survey it is important to make the best use possible of available datasets in different countries, but data that differ with respect to some aspect of the sampling or analytical protocol cannot be combined easily. This subsection considers first the situation in which an estimate of a mean or a total is needed for an area of interest for which data on the target variable are available from two different samples. Then this subsection considers also shortly the situation in which a map is also needed.

These samples can be two probability samples or a non-probability and a probability sample. For example, the area of interest is a country in the European Union and data on the target variable are available from a national sample and a European sample. To obtain an accurate estimate of a mean one wants to combine the national and the European sample. Crowd-sourced data or volunteered geographic data are often collected by non-probability sampling and can be combined with data from probability samples to obtain design-based estimates of means and totals (Laso Bayas et al., 2016; Stehman et al., 2019).

4.3.2.1 Combining two probability samples

If both samples are probability samples, then the inclusion probabilities of the sampling units in both samples can be used to obtain an estimate of a mean or total for the area of interest. Thompson (2002, Section 6.2) describes how to account for unequal inclusion probabilities for finite populations using the Horvitz-Thompson estimator. This estimator can be translated to infinite populations by replacing population size with areal size and probabilities with densities (sampling units per area). If the supports differ between the two samples, then it is necessary to weight the data of the samples accordingly. For a recent discussion on linear combination estimators of population totals and their variances from two independent probability samples we refer to Grafström et al. (2019).

If the two probability samples to be combined have been taken at two occasions, then one might be interested in the change in mean or total from one occasion to the next, in the average mean or total over both occasions or in the mean in the most recent occasion (Cochran, 1977, Section 12.10). Cochran (1977, Section 12.11) provides a composite estimator for the mean in the most recent occasion, see also de Gruijter et al. (2006, Eq. 15.9) in the context of rotational designs. Note that the composite estimator can also be applied to combine two probability samples taken at the same time.

4.3.2.2 Combining a probability sample and a non-probability sample

There are two main reasons to combine a probability sample and a non-probability sample in design-based estimation of means and totals. First, the accuracy of design-based estimates of means and totals obtained from a probability sample can be improved by utilising data that were collected in a non-probability sample. Second, taking a probability sample additional to a non-probability sample enables the use of information from the latter in design-based inference.

Data from a non-probability sample and a probability sample can be combined in several ways to obtain design-based or model assisted estimates of means and totals. Which way will be most attractive depends on factors such as which sample was taken first. We summarize as follows:

1. *Use the non-probability sample to identify strata prior to stratified random sampling* (Brus and de Gruijter, 2003, p. 304). The data from the non-probability sample are used to make a map of the target area by some interpolation method. Next, this map is used to define sub-areas that are used as strata in stratified random sampling in a next stage. For example, Thiessen polygons can be used as strata in nearest neighbour interpolation, or strata can be created by classification of the interpolated values obtained by other interpolation methods.
2. *Create a certainty stratum*, containing all sampling units of the non-probability sample, all selected with inclusion probability equal to one (Stehman et al., 2019, Subsection 4.1). From those units not being part of the certainty sample a probability sample can be taken, if not already available, and this sample can be combined with the certainty stratum in design-based estimation, see Stehman et al. (2019) for details. This approach might be particularly useful in finite populations.
3. *Use a model-assisted estimator*:
    a. A map is constructed using the data from the non-probability sample and next *the post-stratified estimator* is applied (Stehman et al., 2019, Subsection 4.2; Stehman,

2013; de Gruijter et al., 2006, p. 116-117). This approach will particularly be of interest if the non-probability sample was not available for stratification *a priori*.

b. *The difference estimator* (Brus and de Gruijter, 2003, Section 2.1, Eq. 5), based on differences between interpolated data from the non-probability sample and observed data from the probability sample.

c. *The regression estimator* (Brus and de Gruijter, 2003, Section 2.1, Eq. 9; de Gruijter et al., 2006, p. 117-120). In contrast to the difference estimator, the interpolated values are not directly used as proxy values of the target variable, but the estimated slope coefficient of the linear relationship between the true values and the interpolated values in the probability sample is used.

4.3.2.3 Model-based combination of two samples

A general basis for combining soil datasets from different sources using a model-based approach is the linear model of coregionalization (LMCR). This approach allows to model the joint spatial distribution of the different datasets. The LMCR could help to elucidate the effects of the sample support (see Section 4.4), and to study correlation between the underlying variables. Finally, the LMCR can be used to make spatial predictions of the variable on one dataset support by cokriging the different datasets, see Lark et al. (2019) for more details. One drawback of this approach is that the LMCR cannot handle a large number of variables as the approach needs assumption on the multivariate distribution. To estimate the cross-variograms of a LMCR one needs paired observations of both variables that are usually not in general, available at coincident sites. In such cases, the pseudo-cross variogram may be used.

4.4 Other choices in sampling

4.4.1 Sampling protocol

Once aim (e.g. for monitoring purposes) and design (e.g. grid) of a sampling campaign are chosen and agreed between all actors (e.g. ministries, private companies, financial institutions, surveyors) several other considerations need to be decided regarding the way the sampling itself will be done. This includes for example:

- the number of samples to be collected on the field and the area of collection (also called sample support),
- the way samples are collected at the field (e.g. core sampling to keep samples undisturbed or spade/auger sampling),
- the depth of collection (e.g. sampling according to soil horizons by opening soil pits or sampling at the same depth by auger or spade),
- the way samples are managed (e.g. mixed to create composite samples or kept as individual samples). Note that depending on the analysis to be performed different sampling operations may be needed (e.g. sampling for bulk density),
- which standard sampling protocol or field description protocol will be used,

- how samples are stored until analyis; in closed bags either at ambient temperature for physical and chemical analysis or chilled in case of biological and biochemical processes to halt all biological processes occurring in the soil after sampling before analysis
- how and which metadata (data about the sampling itself, the location and surroundings) will be stored.
- How preprocessed soil samples (air dried and sieved to 2 mm) are stored in in archives

That information is needed when comparing and using different datasets. All these possible variations will be analysed and developed in a further deliverable (D6.3 managed by INRAE and due in summer 2021).

4.4.2 Support

Another entity that has not been mentioned yet but which is relevant to consider is the sample or measurement support. The support can be described as the shape and size of a sample unit, the xyz space which is sampled or from which a sensing signal originates. In point soil sampling the support is usually defined as the exact coordinate or an area (point with radius of 2 meter, block, cone shape, field or management zone) with its sampling depth. For instance, the LUCAS Soil sample support is a 4x4 meter block. The soil of a block is typically assessed using a composite sample made up of a set of *n* individual samples; in the case of LUCAS Soil these are five samples. The sample represents the average of the block.

Relevant to consider for a sampling survey is the sampling support that is most appropriate for the objective of the survey and the support of possibly available ancillary information as covariates or previous surveys of the same target variable in the domain of interest. The larger the support, the more short-distance variation (and extremes) will be averaged out. Small blocks (< 10 m) could provide more robust data for mapping and monitoring than points: variation within a sampling unit will be much smaller than the variation between sampling units and variation at extremely short distances is often not relevant for mapping and monitoring purposes. Large blocks will reduce the variation between sampling units. When chosen too large there is a risk that relevant variation, e.g. at landscape scale, is averaged out. Increasing the support size is also at the expense of the precision of the measurement of the block averages (when the number of subsamples is kept constant): five samples within a 5x5 m block will give a more precise estimate of the mean soil property values of that block than five samples within a 100x100 m block.

For proximal and remote sensors the sensor support and therefore the most appropriate sampling support is defined by the measurement method and the specific instrument and its application, also referred to as measurement support. For sensor calibration it might be better to use a small sample support exactly where the sensor measurement is done rather than a larger sample support that would perhaps be used if the purpose was to create a map. At the same time, the sample support should adequately represent the soil within the measurement support to allow proper calibration of the sensor data, either for direct estimation or for use as covariates. For instance for gamma-ray spectrometry the support will be 0.5 m^2 when the sensor is placed directly on the soil, 6 m^2 if the sensor is raised to 40 cm driving height and much more when the sensor is flown on an UAV at 10 m elevation (van der Veeke et al., 2021). It is very relevant to adjust the xy sample support accordingly.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

In all cases the measurement depth will be approximately 30 cm because this depends on the measurement principle and not on the application or height above the soil.

For EMI, electromagnetic induction, measurements the xy support depends on the specific configuration of the sensor, while the z support depends on measurement height.

When using satellite imagery to estimate the soil organic carbon concentration of the soil, it is relevant to realise that the signal measured by the satellite will only be representative of the 0,5 or less centimetres of the soil surface. It is therefore important that the samples are taken at appropriate (very shallow) soil depth to get a good calibration model. In agricultural areas where the soil is ploughed it can be assumed that the soil organic carbon content is homogeneous in the tillage layer and a topsoil sample of the tillage layer at a block that corresponds with an area smaller than the satellite can be used to get a representative sample for calibration.

4.5 Choices within EJP SOIL

As in this chapter on statistical sampling methods becomes clear, there is not one sampling method that fits all possible aims and campaigns. Depending on the purpose(s) of the sampling campaign (estimating a mean, mapping, monitoring, gap filling/additional sampling into an existing scheme) a choice needs to be made on the most appropriate design. In general, we can conclude that to estimate global quantities, such as means and totals, probability sampling approaches with design-based inference are more suitable than model-based methods (Brus & de Gruijter, 1997). For sampling for mapping, model based designs are considered more appropriate, see Brus (2019). In the designs for monitoring, not only spatial variation is a factor, but temporal variation must be taken into account as well. Within the context of EJP SOIL and designing soil monitoring networks with the aim to provide estimates of global quantities in space-time, the P + P and the hybrid P + NP are most relevant approaches to consider, i.e., probability sampling in space.

When choosing a design a general rule of thumb is to keep a design as simple as possible. Overton and Stehman (1995) note that complex sampling designs require complex computational formula for statistical inference. By keeping a design simple, standard statistical analyses commonly used by environmental scientists are correct or at least adequate. Furthermore, they note that the primary concern when designing a monitoring scheme should be to develop an adequate design that makes good use of the available resources and not to construct the perfect, optimal design. Practical convenience and simplicity cannot be sacrificed to achieve optimal statistical efficiency (Overton and Stehman, 1995). On the other hand, practical convenience and simplicity should not be the reasons for cumbersome and complicated statistical inferences.

When the aim is to combine data from two designs by far the most important aspect is to know which designs have been used including the details of the construction of the design, such as which strata were used for instance. When the design and, for probability sampling, the inclusion probabilities of the sampling units are known, this can be used to obtain an estimate of a mean or total for the area of interest.

How to combine national and European monitoring schemes and other aspects of sampling such as metadata storage and sampling protocol will be elaborated more in deliverable D6.3. Morvan et al.

(2008) provide an overview of soil monitoring networks in Europe and suggest options for harmonizing these networks.

5 Harmonised procedures for creation of soil maps

Authors: D. Arrouays, M. Fantappiè, L. Borůvka, C. Piccini, D. Walvoort, B. Stenberg, J. Wetterlind, C. Calzolari, V.L. Mulder, S. Madenoglu, V. Penížek, G. Aust, E. Leitgeb, L. Poggio, R. Skalsky, A.B. Møller, F. van Egmond, F. Ungaro

5.1 Conventional and digital soil mapping

Soil mapping, the process of projecting soil properties and classes in geographic space, is an essential objective of soil science. Soil maps have long been produced with conventional soil mapping (CSM) techniques. CSM has set the conceptual bases for the science of soil mapping, formalized in the work by Jenny (1941). This work states that soils can be described by the main environmental soil forming factors, which are: climate (cl); organisms (o); topography (r); parent material (p); and time (t). Jenny found that soil properties do not vary randomly from place to place. Natural soil bodies are the result of climate and living organisms acting on parent material, with topography or local relief exerting a modifying influence and with time required for soil-forming processes to act. Jenny found that soils are almost the same wherever all the elements of the five factors are the same, i.e. under similar environments in different places, soils are similar. This regularity allows to predict the location of many different soils (Soil Survey Division Staff, 1993). Jenny's factors have been further evolved inside the SCORPAN model (McBratney et al, 2003), which is at the basis of Digital Soil Mapping (DSM).

CSM is expensive and very labour-intensive and often has unquantified uncertainty, thus it is currently mostly used for soil classification mapping and local or regional scale soil property mapping. The availability and accessibility of geographic information systems, global positioning systems, remote sensing (RS) images, digital elevation models (DEM) and terrain parameters, predictive or inference models, and software for data analysis have greatly changed the science of soil survey and mapping and led to the emergence of Digital Soil Mapping (DSM). Using this broad range of data sources and methods, DSM aims to provide up-to-date and accurate maps of soil properties and soil types. DSM is defined as the creation of geographically referenced soil databases based on quantitative relationships between spatially explicit environmental data and measurements (McBratney et al., 2003).

Both conventional and digital soil mapping need soil observations and covariates characterizing the environment where the soil is formed, and a model to derive the information from the input data (Dobos et al., 2006). Conventional soil survey science has set up the scientific base for soil mapping through the identification of the soil forming factors, but the soil-landscape models were qualitatively described (for further details see paragraph 5.3). Conversely, the computer-assisted production of digital maps implies use of mathematical and statistical models combining soil observations with correlated environmental variables and remote sensing images (for further details see paragraph 5.6). DSM techniques can be used not only to map new areas but also to upgrade the quality of previous maps (Grunwald, 2009).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

5.2 Relevant previous and ongoing projects and initiatives

At the seventh conference of the International Society of Soil Science (1960), now the International Union of Soil Science (IUSS), a recommendation for the publication of a global soil map was endorsed. At the congress, several regional maps were presented, including one for Eastern Europe (Hartemink et al., 2013). A European working group on soil was already established at FAO since the early 1950s led by Tavernier, Dudal, Osborne and Moorman, that finally published a Soil Map of Europe in scale 1:2.5 million (Tavernier et al., 1963). A uniform legend was adopted, and various soil classes were later retained in the legend of the FAO-Unesco World Soil Map (FAO-Unesco, 1971–1981). More than 50 soil scientists from 23 countries provided information for this map (Hartemink, 2006) and participated in field correlation activities.

The next European soil map was then produced in the framework of the 1:5 million FAO – Unesco Soil Map of the World for which preparations began in 1961. The European sheets were published in 1981 (FAO-Unesco, 1981). In 1985, a 1:1 million soil map of Europe was published (Commission of the European Communities, 1985). This map was prepared by Tavernier and Dudal, who also produced the 1963 soil map of Europe and the European sheet of the FAO-Unesco soil map (Hartemink, 2008). The 1:1 million map was digitised in 1986 (Platou et al., 1986) and the derived database, called the Soil Geographical Database of the European Community (SGDB, v.1) version 1.0, constituted the first EU harmonized database.

In response to the needs of the MARS Project – Monitoring Agriculture by Remote Sensing (Vossen and Meyer-Roux, 1995) – of DG VI (Directorate General for Agriculture), of having reliable information on the water holding capacities of European soils for input to a model (CGMS) for forecasting the yields of the main agricultural crops throughout the continent, a "Soil and GIS Support Group" was established in 1990 at JRC. In the previous year (1989), a meeting of Heads of European Soil Surveys was held at Silsoe (UK) to review the activities connected with soil survey and data collection throughout Europe (Hodgson, 1991). This led, in 1996, to the establishment of the European Soil Bureau (ESB), as a Project of the Environment Institute (EI) of the Joint Research Centre (JRC) of the European Commission (Montanarella et al., 2005). The ESB continued its activities as European Soil Bureau Network (ESBN) operating through a network of European centres of excellence in Soil Science. The major outcomes of the activities of ESBN were: the Soil Geographical Database of Eurasia at scale 1:1 million (SGDBE), the Pedotransfer Rules Database (PTRDB), the Soil Profile Analytical Database of Europa (SPADE), the Database of Hydraulic Properties of European Soils (HYPRES). In 2005, the Soil Atlas of Europe was published (European Soil Bureau Network of the European Commission, 2005), the first of the Soil Atlases published by JRC. It is worth to be mentioned that most of the soil mapping units boundaries on the maps are inherited from the 1985 CEC map (Hartemink, 2008) witnessing the continuity of the European conventional mapping until the digital era.

All these activities converged into the European Soil Information System (EUSIS), designed to be the main source of georeferenced information on European Soils[176]. The basic structure of EUSIS (then MEUSIS) was finally inherited by ESDAC, the thematic centre for soil related data in Europe hosted at JRC, which is now maintaining and exploiting what was produced in more than 20 years by the

[176] https://esdac.jrc.ec.europa.eu/resource-type/european-soil-database-soil-properties

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

European soil community. The ESDAC ambition is to be the single reference point for and to host all relevant soil data and information at European level. It contains a number of resources that are organized and presented in various ways: datasets, services/applications, maps, documents, events, projects and external links[177] (Panagos et al., 2012).

Under the ESDAC umbrella, all the datasets collected along the years are stored, namely: the Soil Geographical Database of Eurasia at scale 1:1 million (SGDBE); the Pedotransfer Rules Database (PTRDB); the Soil Profile Analytical Database of Europa (SPADE); the Database of Hydraulic Properties of European Soils (HYPRES). In addition, derived data per soil function and threat are available. Also a soil profile database is stored, derived from several years of European soil science activity. The soil profile analytical database for Europe (SPADE), contains standardised soil analytical data across the European countries (Kristensen et al., 2019). Some 1820 soil profiles are now stored in the level-2 database (SPADE 18; Kristensen et al., 2019). The limited numerical consistence of this database should be overcome by the new LUCAS campaign (LUCAS 2018 – SOIL COMPONENT) (Fernández-Ugalde et al., 2017), which should provide measured soil data from 27.000 profiles covering the European area (Kristensen et al., 2019), integrating what is already available for topsoils (LUCAS database, 2015;[178]).

As presented at the recent launch of the new EU Soil Observatory (EUSO)[179], EUSO will integrate and further develop ESDAC by: "Collecting high-resolution, harmonised and quality-assured soil information; Supporting the outcomes of targeted research; Fostering networking, cooperation and partnerships among users of soil data and information; Underpinning policy development through meaningful indicators and assessments". This will be elaborated in the next couple of years.

At a global level, several ongoing initiatives can be mentioned, often involving European partners also active in the beforementioned initiatives.

As a result of the collaboration between FAO, the International Institute for Applied System Analysis (IIASA), ISRIC-World Soil Information, Institute of Soil Science- Chinese Academy of Sciences (ISSCAS), and JRC, the Harmonized World Soil Database (HWSD) combines the existing regional soil information worldwide (SOTER, ESD, Soil Map of China, WISE) with the information contained within the 1:5 million FAO-Unesco Soil Map of the World (Nachtergale et al., 2010, FAO-Unesco 1971-1981). HWSD is a 30 arc-second raster database with over 15 000 different soil mapping units.

5.2.1 Initiatives

5.2.1.1 GlobalSoilMap

The GlobalSoilMap project (Arrouays et al., 2014) aims to produce a fine-grid digital soil map of the world. It is a bottom-up initiative (from country to globe) that was initially launched by a consortium of several institutes, and is now a Working Group of Commission 1.5 'Pedometrics' of the International Union of Soil Sciences (IUSS). Maps and data are released on two spatial entities. A primary spatial entity – point support – a pedon located at the centre point of a grid of 3 arc-seconds by 3 arc-seconds.

---

177 http://esdac.jrc.ec.europa.eu
178 https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data
179 https://ec.europa.eu/jrc/en/eu-soil-observatory

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

A secondary spatial entity - a voxel - 100m by 100m horizontal and 0 to 2m vertical dimensions, centred on the point support. Twelve basic soil properties are delivered as predictions with uncertainty at six standard depths (down to a maximum of 2 m: 0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm, 100-200 cm).

Additional soil properties may be predicted at the discretion of the countries but these are not mandatory. There are four tiers related to increasing levels of precision of uncertainty estimates. For tier 1 and tier 2, the estimate of uncertainty is the 90% Prediction Interval (PI). Most advanced countries in the world are Australia and the US. Many other countries delivered partial GlobalSoilMap products, either for part of their geographical coverage or part of the properties. In Europe, the main advanced countries or sub-countries are France, Scotland, Denmark, and Hungary (e.g., Mulder et al., 2016; Poggio and Gimona, 2017; Adhikari et al., 2014; Pasztor et al., 2020) The GlobalSoilMap WG is also member of the Pillar 4 working Group of the UN-FAO Global Soil Partnership[180].

5.2.1.2 Global Soil Partnership

The Global Soil Partnership (GSP) was established in December 2012 at FAO with the aim of enhancing the collaboration among all soil-related stakeholders, from academia to policy makers to farmers. One of the key objectives of the GSP is to improve the governance and promote sustainable management of soils. Among the pillars of action of the GSP, Pillar 4 has the mandate to "Enhance the quantity and quality of soil data and information". (chapter 1.2.7 and 3.1)

The major objective of Pillar 4 is to develop a spatial data infrastructure (GloSIS[181]) that brings together soil information collected by national institutions (see Chapter 3). To this aim, the International Network of Soil Information Institutes (INSII) has been established, which represents the backbone for implementing the plan of action of Pillar4. Following the decisions taken by the GSP Plenary Assembly, that is the main governing body of the partnership, global map products are planned to respond to topical issues. These will use data from and can be incorporated in GloSIS. In 2017 the Global Soil Organic Carbon map (GSOC map) was produced (FAO & ITPS, 2020); the Global Soil Salinity Map and the Global Soil Organic Carbon sequestration potential maps are constructed now and will be released in June 2021; the Global Soil Erosion Map is planned for release in 2021-2022. These maps are developed following the general GSP principle of being a country-driven initiative.

5.2.1.3 SoilGrids

SoilGrids[182] is a system for global digital soil mapping that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe facilitated by ISRIC-World Soil Information (chapter 3.1.1.5). SoilGrids prediction models are fitted using over 230 000 soil profile observations from the WoSIS database (Batjes et al., 2020) and a series of environmental covariates. Covariates were selected from a pool of over 400 environmental layers from Earth observation derived products and other environmental information including climate, land cover and terrain morphology.

[180] https://www.isric.org/projects/globalsoilmapnet

[181] http://www.fao.org/global-soil-partnership/areas-of-work/soil-information-and-data/en/

[182] https://soilgrids.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

The outputs of SoilGrids are global soil property maps at six standard depth intervals (according to the GlobalSoilMap IUSS Working Group and its specifications, see above) at a spatial resolution of 250 meters. Prediction uncertainty is quantified by the lower and upper limits of a 90% prediction interval. Maps of the following soil properties are currently available: pH, soil organic carbon content, bulk density, coarse fragments content, sand content, silt content, clay content, cation exchange capacity (CEC), total nitrogen, soil organic carbon density and soil organic carbon stock. Reference for the latest version of SoilGrids is: de Sousa et al., 2021 (in review).

5.2.2 Projects

5.2.2.1 ESA WorldSoils (Copernicus Land Monitoring Service)

WorldSoils is part of the European Space Agency's (ESA) Earth Observation Strategy 2040 and is ESA's Earth Observation Envelope Programme backbone[183]. As part of the Earth Observation Strategy 2040, ESA includes a "science for society" element in each area of the strategy. This component of the strategy is achieved by work, like WorldSoils, which transfers scientifically proven Earth observation research results into pre-operational products that meet the most important needs of user organisations and public authorities.

The project's methods are based in Earth Observation (EO) technology with the main objectives of:

1. Developing a pre-operational global Earth Observation-Soil Monitoring System based on monitoring topsoil organic carbon (SOC). This will integrate spectral modelling with DSM approaches.

2. Engaging and bringing together authoritative end users for developing soil indices relevant for monitoring global topsoils.

5.2.2.2 eSOTER

Within the Framework of Global Earth Observing System, the eSOTER project[184] (2008-2012) addressed the need for a global soil and terrain database, essential for many interpretations in the field of agriculture, environment, watershed management, infrastructure, etc. The collaborative project (14 partners, ISRIC - World Soil Information as coordinator) focused on the development of various products overcoming the major barriers to a comprehensive soil observing system: i) morphometric descriptions in SOTER DEM methodology and in newly developed DEM analysis; ii) characterization of soil parent material and soil pattern using remote sensing techniques; iii) recognition of soil pattern using existing legacy data converted into a standardized SOTER format. The analyses were performed in four selected window areas at scale 1:1 million and four pilot areas at scale 1:250.000. The final result of the eSOTER project was a Pilot Platform and a portal that provides, among others, open access to i) a methodology to create 1:1 million SOTER for four windows, ii) an artefact-free 90m DEM for the windows; iii) methodologies to create 1:250.000 scale enhanced SOTER databases and the databases themselves for the pilots, iv) advanced RS techniques to obtain soil attribute data and v) applications

183 https://www.isric.org/projects/esa-worldsoils
184 https://www.esoter.net/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

related to some major threats (e.g. soil erosion and soil compaction) to soil quality. The results have been disseminated through stake-holder conferences and web-based services.

5.2.2.3 DigiSoil

The European collaborative project of the 7th Frame Programme (FP7) Digisoil - Integrated system of data collection technologies for mapping soil properties - was carried out between 2008 and 2011 (Grandjean et al., 2010)[185]. The consortium included 10 partners from 7 countries. The main objective was to integrate and improve in situ and proximal measurement technologies for the assessment of soil properties and soil degradation indicators, going from sensing technologies to their integration and application in (digital) soil mapping (DSM). Specific attention was paid to the feasibility of such developments based on economical constraints, reliability of the results and needs of the DSM community. Similarly to iSOIL, this project also brought advancement in the methods of soil and auxiliary data collection including the standardization of processes and technical specifications in terms of equipment (sensors, acquisition system, mobile vector), techniques (signal processing, inversion or fusion processes, specialization) and operational protocols.

5.2.2.4 iSOIL

The European collaborative project of the 7th Frame Programme (FP7) iSOIL – Interactions between soil related sciences – Linking geophysics, soil science and digital soil mapping was carried out in the years 2008-2011 (Werban et al., 2010; [186]). The consortium included 20 institutions and companies from nine countries, the coordinator was UFZ Helmholtz Centre for Environmental Research Leipzig, Germany. The project did not aim to create large databases or maps, but it focused rather on the methodological approaches. The objectives of iSOIL research were the development of new and the improvement of existing methods that included geophysical, spectroscopic and monitoring techniques. It was based on the integration of three major components: (i) high resolution, non-destructive geophysical (e.g. Electromagnetic Induction EMI; Ground Penetrating Radar, GPR; magnetics, seismics) and spectroscopic (e.g., Near Infrared, NIR) methods, (ii) concepts of Digital Soil Mapping (DSM) and pedometrics, as well as (iii) optimized soil sampling using profound soil scientific and (geo)statistical strategies. The project yielded a number of papers and methodologies that can be used in collecting both soil and auxiliary data for digital soil mapping and assessment.

5.3 Soil data

The main input data for soil mapping are soil data, that is, any type of data originating from soil. Soil data are used as input variables either for CSM and for calibrating DSM prediction models. They are also used as ground truth for statistical validation of the map produced. Either CSM and DSM need as an input also ancillary variables, representative of the soil forming factors. Ancillary data, or

---

[185] http://digisoil.brgm.fr/default.aspx

[186] https://esdac.jrc.ec.europa.eu/projects/isoil

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

environmental covariates, are data showing relationships with some soil properties and available over the coverage of interest. These are described in chapter 5.4.

5.3.1 Point soil observations

Point soil data may come from soil profiles (pits), soil augering, and various types of other samplings and information. At the global level, data are spatially irregularly distributed, with some parts of the world being relatively densely surveyed (e.g., Europe, U.S.A.) while other parts having still very sparse point data or no data at all. Furthermore, the soil data accessibility is often restricted. A bottom-up participatory mapping procedure, as adopted by the Global Soil Partnership, can overcome this problem, which in general can be addressed through the establishment of sharing agreements with the soil data owners. There are large discrepancies between countries, either in the total number of soil profiles compiled or in the efforts put in place in data rescuing (Arrouays et al., 2017). Recently, ISRIC – World Soil Information used its WoSIS database containing ca. 240.000 points for version 2.0 of SoilGrids (de Sousa et al., 2021). Typically, point data are often more abundant in surface layers (topsoils) than in deeper layers. They often include a profile description, and soil classes according to national or international classification systems, and various related analytical properties. In some cases, it may even include *in situ* measurements, such as for instance proximal sensing (PS) in the field, or hydraulic properties measurements such as infiltration rates, and so on. In some cases, soil data are collected using fixed depths without including a soil profile description (for instance in EU-LUCAS, Tóth et al., 2013). Another inconsistency between countries or even between different surveys within countries is caused by different analytical methods used (chapter 3.4). . In most of the countries, the majority of centralized point data are legacy data that were collected in the past for various purposes (mainly conventional soil mapping and agriculture) using various designs (chapter 4) and then harmonized. Thus, the density of data varies a lot, even inside one country, and some data may be rather old, from the 1950s' to now. More frequently analyzed soil properties are often used to derive pedo-transfer functions (chapter 3.4) in order to estimate soil properties that are scarcely available in databases (e.g., bulk density, available water capacity). Various proximal soil sensing products can be also used as alternative or surrogate to soil observations for calibrating the DSM models (Adamchuk et al., 2004; Lagacherie and Gomez, 2018).

5.3.2 Soil maps

Soil properties can be directly derived from legacy soil maps when their polygons include information on the spatial coverage of soil types, and the soil types are characterized by quantified attributes, including for instance soil classification, mean and standard deviation of soil properties or modal values of categorical soil properties by site description (e.g. soil rooting depth, soil drainage), for the soil type as a whole or given by layer depth or horizon. This can be obtained by rasterisation of detailed soil maps and a computation of weighted means (Odgers et al., 2012) and estimation of prediction intervals (Helmick et al., 2014), or a disaggregation of less detailed soil maps using tools such as DSMART (Odgers et al., 2014). Another use of soil maps in DSM is to use them as *reference areas* where there is a detailed understanding of soil distribution and its controlling factors, to form a basis for extrapolation to a broader domain (Lagacherie et al. 2001; Grinand et al., 2008).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

5.4 Ancillary data, environmental covariates

Soil mapping, either by CSM and by DSM, needs ancillary information. These ancillary information are environmental covariates which have a presumed relation with the soil forming factors, firstly described by Jenny (1941), which evolution is the SCORPAN model (McBratney et al., 2003; Wadoux et al., 2020). They may also show relationships with some soil properties, without being *stricto sensu* the controlling factors of soil formation. Their use depends on their availability, their full coverage of the area of interest, predictive power, and pedological relevance. In order to avoid over-fitting, the selection of covariates should respect the following criteria: parsimony, non multi-collinearity, pedological relevance, and relative importance (Arrouays et al., 2020; Wadoux et al., 2020; de Sousa et al., 2021). Inclusion of non-informative covariates increases model uncertainty, particularly for linear models. Covariate reduction (also known as feature selection) is also important because as the number of covariates increases so does the chance of model overfitting and the amount of computation time (de Sousa et al.,2020). Moreover, simpler models are easier to interpret.

Semi-automated covariate selection methods can be grouped into two broad categories: unsupervised and supervised. Unsupervised methods evaluate covariate relevance outside of a predictive model by selecting covariates that pass some criterion. Supervised methods select optimal covariates by identifying the covariate set that maximizes model predictive ability (Kuhn and Johnson, 2013). Unsupervised covariate selection methods include correlation analysis, Optimal Index Factor (OIF), and principal components analysis (PCA).

5.4.1 Covariates related to soil properties (S factor)

Some covariates, mainly derived from Proximal Sensing (PS) or Remote Sensing (RS) products, are often used, not necessarily because they have a causal relationship to soil formation, but because they may be correlated to soil properties. This is the case for numerous PS and RS data, particularly on bare soils and derived indices which can be used to infer relationships with parameters of interest such as soil organic carbon, carbonate content, soil texture, soil moisture, soil Fe content, soil salinity, etc. (see reviews by Mulder et al., 2011; Viscarra-Rossel et al., 2006; Adamchuk et al., 2018). Proximal soil sensing (handheld, driving, UAV) can provide detailed information on within-field or farm soil variation that are useful for DSM for precision agriculture management, but are more difficult to use for mapping of larger geographical areas. Remote sensing (aerial, satellite) on the other hand often provides larger geographical coverage but may have lower spatial resolution. A large number of spectral indices applied to bare soils have been used in DSM studies (Boettinger et al., 2008; Mulder et al., 2011; Mahmoudabadi et al., 2017). A promising covariate is derived from remote sensing radar data (such as Sentinel 1 and PALSAR), see for example Poggio and Gimona (2017). Gamma radiometric data are related to the composition of minerals and the intensity of weathering and illuviation resulting in different soil textures (Wiltord, 2012), and can capture variations even under vegetated soils up to about 30 cm depth (Martelet et al., 2013). RS hyper-spectral data are more and more available and used (see a review from Lagacherie and Gomez, 2018). Most of these data, however, mainly come from the reflectance or the emission of topsoil (top centimetres), and their efficiency decreases greatly with soil depth. Some deeper soil properties (e.g. rooting depth, available water

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

capacity) may, however, be related to the response of vegetation, such as NDVI during dry periods, or thermal infrared imagery, which has been recently used as a dynamic covariate for salinity (Ivushkin et al., 2019). Magnetometry, mainly measuring particularly magnetic susceptibility, is a PS technique that assesses the concentration of strongly magnetic iron-containing minerals in the soil (Thompson and Oldfield, 1986; Evans and Heller, 2003). It can reflect soil microenvironmental conditions (Cornell and Schwertmann, 1996) like oxidative/reducing environment, temperature, humidity, etc. Magnetometry is often used also for the indication of soil pollution with potentially toxic elements like lead as the magnetic minerals can be of anthropogenic origin coming with dust from various industrial emissions (Petrovsky and Elwood, 1998). Other PS sensing techniques that are used as covariates or proxies for soil properties are Electomagnetic Induction (EMI) and Electrical Conductivity (EC). Both reflect differences in a combination of moisture, texture, salts concentration and porosity up to 1 to 4 meters depth. Ground Penetrating Radar provides insight in changes in texture and moisture (di-electric constant) in the soil profile with a relatively high resolution in depth up to 2-to-6 meter depth. When used stand-alone all proximal sensing techniques described here need calibration by soil samples or profile descriptions and are often applied at field or farm scale. A comprehensive overview of a number of PS techniques is given by Adamchuk et al. (2018)

Legacy soil maps produced by conventional soil mapping (Section 5.5) delineate polygons inside which uniform soil forming factors are recognized, so that the internal variability of soil properties and soil types inside each polygon is smaller than the variability in-between polygons. Therefore, legacy soil maps may be used as categorical covariates in DSM (McBratney et al., 2003), and their predictivity has been demonstrated (Fantappiè et al., 2010, 2011).

5.4.2 Climate (C factor)

Climate typically does not have short scale spatial variability like other factors do. Covariates related to climate mainly include temperature and precipitation time series. The dimensionality of these covariates often has to be reduced. Other climatic co-variates include potential evapotranspiration, snowfall, cloud cover, solar radiation, water vapour, wind speed, etc., or more aggregated products such as bioclimatic zones, which incorporate information both on climate and vegetation. It is important to keep in mind that these different covariates can be combined in various indices, such as bioclimatic variables (Waltari et al., 2014; Fick and Hijmans, 2017; Poggio et al., 2018; Chen et al., 2019a).

5.4.3 Organisms (O factor)

Organisms are mainly represented by biomes and vegetation: land use/land cover classes derived from RS and/or from historical maps, potential vegetation, vegetation indices (e.g., NDVI, NPP) derived from RS bands of various satellites (Landsat, MODIS, Sentinel, etc.). Further research is needed to assess if high spatial and temporal resolution products (such as Sentinel 2) allow better characterization of agricultural practices, vegetation cover during crop rotations, crop residues and bare soil periods. Thus, the influence of some human activities could be incorporated in modelling. However, most of the socioeconomic data related to human behaviour are generally not part of the paradigm. Soil biodiversity and activity are not often used as we lack the data, but these can be important factors of

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

soil formation. For the availability of mapped data on soil management among EJP SOIL countries see chapter 2.6.2 and annex 2. In CSM the O factor was considered by aerial photograph interpretation by delineating areas homogeneous for land use.

5.4.4 Relief (R factor)

These covariates are used the most in DSM, at least when the terrain is not completely flat. This is due to the following reasons: i) the large availability of various Digital Elevation Models (DEM) at different resolutions, ii) the effect of relief on water flows and accumulations, and its major influence on soil erosion, transportation and deposition processes, and iii) the elevation, at least at broad scale, is generally correlated with climate and vegetation. A very large number of topographic indices derived from DEMs are available in the literature (e.g. Florinsky, 2012; Hengl et al., 2017; Wilson and Gallant, 2000). It is important to consider that topographic indices may reflect different processes and controlling factors, depending on the size of the neighbourhood used to compute them. Thus, a multi-scale approach is sometimes used, by combining different neighbourhoods (Grinand et al., 2008; Behrens et al., 2018a). Landform classes can also be derived from DEM and used as covariates Various DEM, their sources and uses are described in Hengl et al. (2017). In CSM the R factor was considered by stereoscopic studies aimed at the delineation of landforms.

5.4.5 Parent material (P factor)

Parent material covariates mostly derive from existing geological or lithological maps (Gray et al., 2016), either in CSM and DSM. These maps often need to be re-classified in order to get relevant information about the conditions of weathering and pedogenesis (Bakacsi et al. 2014; Vaysse and Lagacherie, 2015; Gray et al., 2016). The resolution of these maps is often rather coarse. Another possibility is the use of gamma radiation measurements as they reflect provenance of parent material, either as airborne or proximally sensed data (Loiseau et al., 2020).

5.4.6 Age (A factor)

In soil science the age factor has been studied by the use of soil chronosequences (Huggett, 1998), with the aim to study soil pedogenetic processes. At the base of the construction of a chrono sequence there is the assumption that it is possible to find a geographical or vertical sequence of soils, where all the other factors of soil formation, except of time, are fixed. This condition is quite rare and difficult to assure in the reality. Despite of that, there is a large body of literature available on this topic. Furthermore, the construction of a chronosequence needs the dating of parent materials. When the soil has developed only from its underlying parent material, its age can be estimated using geological or lithological maps. In many cases, however, soil development is the result of the weathering and pedogenesis of both allochthonous and autochthonous materials.

When soil properties are evolving rather quickly, it is possible to insert the age factor inside DSM studies, considering the soil sampling time (De Gruijter et al., 2006, Fantappiè et al., 2010 and 2011; Heuvelink et al., 2006, and 2013; Webster and Heuvelink, 2006), producing a so-called spatio-temporal mapping. In this way, it is possible to evaluate the age effect on the soil properties mapped. Finally, age is the least used covariate in broad scale DSM studies (McBratney et al., 2003; Grunwald, 2011).

5.4.7 Geographical coordinates (N factor)

The spatial configuration of the sampling locations forms the basis of DSM studies based on interpolation by kriging. Kriging of the residuals of a non-spatial model may help to detect trends not well captured by the non-spatial model, which may indicate that a controlling covariate is missing. There is still a debate about incorporating coordinates in pure machine learning (ML) models. Wadoux et al. (2020) recently questioned the use of coordinates, or the inclusion of various distances (Hengl et al., 2018; Behrens et al., 2018b), as covariates in ML models. They argued, among other, that: “it precludes analysis of the residuals and the generation of new hypotheses from these residuals” […] or “may well integrate over several pedologically relevant covariates, making them better covariates or masking the effect of pedologically relevant covariates”. In practical terms, the inclusion of coordinates and distances has yielded higher accuracies and meaningful information in some cases (Møller et al., 2020), whereas in other cases, it proved to be misleading (Meyer et al., 2019). It is likely that the usefulness of geographic coordinates depends on the specific setup, and researchers should therefore be cautious if they decide to include them in ML models.

5.5 Conventional soil mapping

5.5.1 Introduction

In conventional soil mapping (CSM) (Minasny and McBratney, 2016) soils are reproduced on maps based on a soil surveyor's conceptual or mental model (Hudson, 1992), verified with field observations (Legros, 2006). CSM typically employs expert-based survey methods to create soil maps. First, a mental soil-landscape model is made. Soil boundaries are defined based on landscape features from aerial photograph/DEM interpretations. Next, sample locations are selected that are likely to be most informative, and their spatial position is optimized by the surveyor in order to increase their efficiency. The mental model is refined based on these field observations. Additional samples might be obtained and finally the map unit composition is determined. The map is a general-purpose map with soil classes and additional soil profile descriptions characterizing each map unit (Bregt, 1992). Central to CSM is the description of the soil by a soil classification system (IUSS Working group WRB., 2006).

5.5.2 Conventional soil mapping process

CSM involves delineating segments of the landscape with similar soil characteristics or classes. Surveyors observe purposive positions in the landscape to cover variability in soil forming factors. The actual placement of the observation points is driven by surveyor decision.

In the first step, distinctive landscapes are outlined mainly by interpretation of aerial photographs, topographic maps, geologic maps, land-use maps, climatic-maps, remotely sensed data, and other available covariate information (chapter 5.4). Relief is perceived by stereoscopic study aimed at identifying main landforms (e.g., terraces, flood plains, sand dunes, kames, and eskers). The soil scientist uses geomorphological studies and knowledge as a base to take full advantage of photo interpretation. The mapper develops hypotheses of the soils present in each delineation. The mapper tests those hypotheses by soil augering, soil profiles observations, or observing natural exposures and confirms or rejects each hypothesis.

Traverses/transects of the preliminary delineations are performed and positioned to encounter the maximum variation in each delineation. They are used to determine the composition and design of map units. In the traverses the observation points are taken at any distance apart, and chosen to represent specific areas on the landform, which are supposed to have different soil types. Traverses are used to locate profile locations. Soil profiles are fully described, sampled and observations are made to determine the soil classification.

In the transects the observation have fixed length intervals between observation points, and they are sampled using a soil auger. Systematic variation is quantified and more easily understood with transects. The sampling design obtained by this traversing and transecting process, inside each polygon delineation, results in stratified, purposive sampling (chapter 4). Highly heterogeneous areas are sampled in higher density than homogenous areas where less sampling points are needed (Balkovič et al., 2013). The overall sampling/observation density is connected to the general resolution of the resulting maps. The sampling/observation density is defined for given map scale (Soil Survey Division Staff, 1993).

Field observations confirm or reject the soil-landscape model developed at the first step. The soil mapper modifies the mapping delineations accordingly. At the same time, the soil mapper elaborates the map legend, through the so-called process of soil correlation. Correlation is the process of comparison and grouping of individual soils with similar properties and same taxonomic class, but located in different areas. The correlation leads to the production of soil types and *DerivedSoilProfiles.* A *DerivedSoilProfile* is a soil type, also called Soil Typological Unit (STU), with associated mean and modal values for soil properties (chapter 3.4). A *DerivedSoilProfile* can be seen as a characterisation of a STU, or Series, as recognized in the European Soil Geographical Database and other soil databases at national or regional levels. A broad list of literature describing the soil survey and mapping at national or international level exists (Soil Survey Division Staff, 1993, 2017; Němeček et al., 1967).

5.5.3 Conventional maps

The result of CSM are polygonal soil maps, where polygons represent mapping units, referred to as *SoilBody* (chapter 3.4). The aim and the resolution of the map and the variation present in the mapped area determine the classification level that can be displayed on the map. The maps are commonly general-purpose maps with soil classes (Bregt, 1992). Often the map units also specify other information such as terrain attributes/units or geological attributes. Their definition is based on the purpose and resolution of the soil survey. On large-scale maps, the mapping units correspond to soil classification units. Mapping units of medium and small-scale maps are usually built by combination of several soil classification units because of impossibility of spatial resolution of the single classification units at given scale. Soil associations, consociations and other types of combinations are recognized (Soil Survey Division Staff, 1993, 2017). They vary by spatial distribution and level of heterogeneity of the classification units within the mapping units. Polygonal base maps describe the variability within the mapping units, but few times also give estimated values of extent coverage of the different soil types. The variability can vary between low level (such as in case of large-scale maps) to high extent (for soil associations). The complexity of the resulting mapping units is described by an attribute database.

Together with polygonal maps, databases of *ObservedSoilProfiles* and *DerivedSoilProfiles* are recorded and stored (Tavares-Wahren et al., 2015). These databases serve as the attribute characterization for the mapping units. The databases usually comprise both morphological ad analytical data. The characterization of the mapping units can be done: by single observation (one-to-one connection between observation and polygon),by statistical set of points characterizing the mapping units (many-to-one), by so called representative profiles, that is by the choice of the most typical profile (central concept), or with the elaboration of *DerivedSoilProfiles*.

Conventional soil maps are accompanied by a soil handbook, which contains interpretations and general sections covering such topics as climate, physiography, relief, drainage, geology, and vegetation, which relate to the soil types in the area. These characteristics improve the understanding of the properties, distribution, use, and management of the soils.

5.5.4 Past and present of CSM

CSM has a long tradition in Europe and other regions. Most of the soil maps of European countries are based on CSM (Panagos et al., 2011). National based soil maps serve for production of higher scale maps such as the Soil Map of Europe 1:1M (ESBN, 2005; EC and ESBN, 2004). Different approaches, aims, concepts of soil surveying and mapping in countries cause problems with collating of these sources to such general maps and results in trans-boundary discrepancies (Thompson et al., 2012). These problems can be overcome by approaches described more in detail in chapter 5.8.

When exploiting conventional soil maps, the users must take into consideration certain limitations (Helmick et al., 2014). CSM map contain information on uncertainty when mapping unites are described by *DerivedSoilProfiles* and an estimation of their areal cover inside each soil mapping unit is given. The accuracy of a maps can be assessed by validation, for which an additional sample is needed, preferably collected by probability sampling (see Chapter 4). The data presented in CSM maps can be out of date because of the mapping period. Many of the national based surveys in Europe were done between 1960s-1980s (Finke et al., 1998; Jones et al., 2005).

Nowadays, CSM is replaced on the continental scale mapping and also at several national scale mapping surveys, by approaches based on DSM. DSM techniques are also used now for disaggregating CSM maps. The conventional soil maps are used as well within the DSM approach. 5.6.3.2CSM is still in use at national scale, for example for land taxation in Czechia (Novotný et al., 2013), and subnational scale.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

5.6 DSM techniques

5.6.1 Introduction

The basis of DSM is the application of pedometric methods that predict the spatial and temporal distribution of soil types and soil properties (Figure 5.1). The conceptual framework in which the pedometric methods are applied is the SCORPAN model (McBratney et al., 2003, Section 5.1). A further evolution of the SCORPAN model is the STEP-AWBH model (Grunwald, 2011) which incorporates the soil-ecosystem evolution and anthropogenic forces into the modelling process. DSM relies on field, laboratory and RS and eventual PS soil observations, integrated with quantitative methods to infer spatial patterns of soils across various spatial and temporal scales (Grunwald, 2010), and in correlation with auxiliary data (figure 5.1).

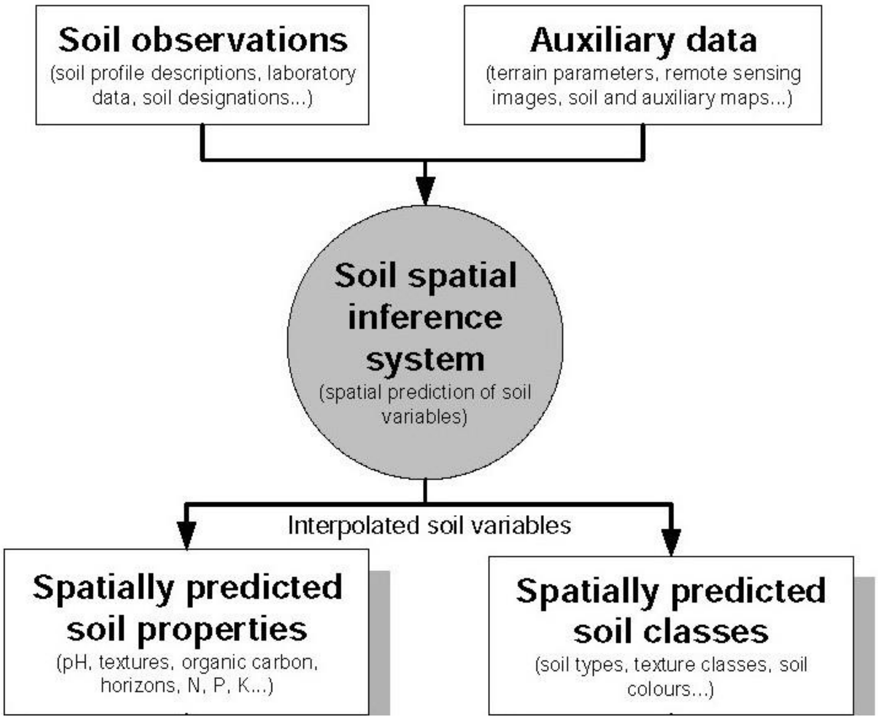


Figure 5.1 Dobos et al., 2006

Typically, the auxiliary data or covariates are exhaustive georeferenced data layers, including digitized geological and soil maps, satellite images and derivatives of the latter (Section 5.4). Using a broad range of data sources and methods, DSM aims to provide accurate soil maps. The DSM approach focuses on prediction of the spatial distribution of soil properties rather than only the soil type. Moreover, it is easy to run a DSM model again when new soil (or covariate) data become available, promptly updating the maps. This makes DSM flexible and more suitable in providing soil information for specific applications compared to CSM.

DSM approaches generally fall in one of the following categories:

- Deterministic interpolation (for example nearest neighbour and inverse distance)
- Geostatistical mapping (for example kriging)
- Spatial statistics (including machine learning)

- Combined approaches

Deterministic interpolators do generally not make use of auxiliary variables. Geostatistical interpolators may or may not use auxiliary variables. Spatial statistics methods require auxiliary variables. The techniques are briefly discussed below but extensive overviews can be found in:

- A review of spatial interpolation methods for environmental scientists[187]
- A practical guide to geostatistical mapping[188]
- A disposition of interpolation techniques (Knotters et al., 2010).
- Book on Pedometrics[189]
- Using R for DSM[190]
- Predictive soil mapping with R[191]
- GSOC Soil Organic Carbon Mapping Cookbook[192]
- An introduction to statistical learning[193]
- Soil spectral Inference with R[194]
- Basic steps in Geostatistics[195]

5.6.2 Methods

5.6.2.1 Deterministic interpolation

Spatial interpolation is the process of predicting the value of a target variable at unsampled sites from available measurements made at point locations within the same area or region (Burrough and McDonnell, 1998). Basically, deterministic interpolators create surfaces from the values of known data points within the study region, based on either the extent of similarity (e.g. inverse distance weighted, IDW) or the degree of smoothing (e.g. radial basis functions, RBF). The first method gives weight to data points such that their influence on prediction is reduced as distance from the point increases, while the second is a real-valued function whose value depends only on the distance from the origin of each RBF.

Deterministic interpolation techniques do not provide any assessment of the interpolation error. They can be divided into two groups, global and local. Global techniques calculate predictions using the entire dataset. Local techniques calculate predictions from the measured points within neighbourhoods, which are smaller spatial areas within the larger study area. A deterministic interpolation can either force the resulting surface to pass through the data values or not. An interpolation technique that predicts a value that is identical to the measured value at a sampled location is known as an exact interpolator. The latter can be used to avoid sharp peaks or troughs in

[187] https://data.gov.au/data/dataset/a-review-of-spatial-interpolation-methods-for-environmental-scientists
[188] http://spatial-analyst.net/book/system/files/Hengl_2009_GEOSTATe2c1w.pdf
[189] https://www.springer.com/gp/book/9783319634371
[190] https://www.springer.com/gp/book/9783319443256
[191] https://soilmapper.org/
[192] http://www.fao.org/documents/card/en/c/I8895EN/
[193] https://link.springer.com/book/10.1007/978-1-4614-7138-7
[194] https://www.springer.com/gp/book/9783030648954
[195] https://www.springer.com/gp/book/9783319158648

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

the output surface. IDW, nearest neighbour interpolation, and RBF are exact interpolators, while global polynomial, local polynomial, kernel interpolation with barriers, and diffusion interpolation with barriers are inexact ones.

5.6.2.2 Geostatistical mapping

Geostatistical interpolation techniques (kriging) utilize the statistical properties of the geo-referenced measurements, quantify the spatial autocorrelation among measurements and account for the spatial configuration of the sample points around the prediction location. The use of geostatistical models has several advantages compared to using deterministic interpolation or empirical models[196]. In addition to the structural part (drift or trend) that is modelled by the empirical models, the spatially correlated random part of variation is modelled. Here, the structural part can be estimated using covariates. Modelling both the structural part and the random part of variation generally results in a higher prediction accuracy; kriging provides the best linear unbiased estimator of an unknown location along with the prediction uncertainty (Goovaerts, 1997). Overall, geostatistical methods allow for more in-depth analysis of prediction uncertainties and spatial processes compared with statistical methods (Chilés and Delfiner, 1999). Despite the many pros for using geostatistical methods some major limitations exist. Calibration of geostatistical models, generally requires higher sampling densities and spatial dependence of the observations (Hengl et al., 2003), modeling non-linear relations between a soil property and covariates is not straightforward, models are computationally demanding, etc. (Wadoux et al., 2020).

Kriging

The term “kriging” depicts a set of spatial interpolation methods that estimate a regionalized variable at selected grid points, predicting values without bias, and with minimum error variance. Kriging relies on the knowledge of the spatial structure of the data, consisting in a form of weighted averaging in which the unknown value in a point or block is predicted from the other known values (Heuvelink and Webster, 2001). The weights are calculated in such a way that points nearby to the location of interest are given more weight than those farther away, and clusters weight less heavily than single points. Thus, the estimator is unbiased.

Kriging can be described as a two-step process: first, the spatial covariance structure of the sampled points is determined by fitting a variogram; and second, weights derived from this covariance structure are used to interpolate values for unsampled points or blocks across the spatial field. The value of each point is calculated in such a way as to minimize the expected error for that point.

A semivariogram is a visual depiction of the covariance between each pair of points in the sampled data. For each pair of points, the gamma-value or semi-variance (a measure of the half mean-squared difference between their values) is plotted against the distance, or “lag”, between them. The experimental semivariogram is the plot of observed values, while the theoretical or model semivariogram is the conditionally negative semidefinite model that best fits the data. Semivariogram models are selected from a limited number of permissible functions, including the Spherical model,

[196] Empirical Interpolation Methods in literature, e.g. Lagrange Polynomial

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

the Matérn model (with the Gaussian and Exponential models as special cases), and the Power model (with the linear model as a special case). Kriging is an "optimal linear predictor" and an exact interpolator (provided that measurement error is ignored), meaning that each interpolated value is calculated to minimize the prediction error variance and set bias to zero for that point. Thus, it is known as Best Linear Unbiased Predictor (BLUP).

Kriging has many useful properties:

- The interpolated value is the most accurate predictor among all linear models.
- It quantifies the interpolation error by means of the kriging variance.
- The measurement error and short distance variation is incorporated in the nugget of the semivariogram.
- It incorporates spatial correlation according to "everything is related to everything else, but near things are more related than distant things". This is often referred to as "Tobler's first law of geography" (Tobler, 1970). Although this is the case for many spatial interpolation methods, in kriging the weights are derived from the degree of spatial correlation, rather than on Euclidean distance only.

The quality of prediction and the kriging properties of unbiasedness and minimum estimation variance can be tested by cross-validation, this involves deleting each sample in turn and then kriging it independently from all other points, or independent validation. In addition, the kriging procedure produces the prediction error variance. See Section 5.7 for more details.

Considering the support (chapter 4.4.2) of prediction, we can distinguish point kriging, applied to areas or volumes of the same size as that of the original sampling unit (*i.e.*, no change of support), block kriging, applied to areas or volumes that are larger than the units that were originally sampled, area-to-point kriging (Kyriakidis, 2004) where the data pertain to areas and the predictions to points, and area-to-area kriging, where both the data and the prediction locations pertain to areas. (Walvoort, 2019).

The most commonly used Kriging algorithms are:

- Ordinary Kriging, by far the most common, is one of the simplest forms of kriging. The weights are quantified through the covariance or semivariogram function.
- Cokriging, which uses the distributions of a second, third, or more, correlated variables (covariates) along with the primary variable to provide predictions. Cokriging can improve estimates if the primary variable is difficult, impossible, or expensive to measure, and the cheaper secondary variables are sampled more intensely than the primary variable (Myers, 1982).
- Universal Kriging, used on data with a significant spatial trend, allows the mean of the values to change in different locations, and the trend is modelled as a function of coordinates only.
- Kriging with external drift, a variant of Universal Kriging in which the trend (drift) is defined externally through some auxiliary variables. Both trend and residuals are modelled simultaneously, and prediction equations are derived using this single model.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

- Regression Kriging, which combines a regression of the dependent variable on auxiliary variables with kriging of the regression residuals (Knotters et al., 1995; Odeh et al., 1995; Hengl et al., 2003, 2007). First the trend is modelled (while ignoring spatial correlation in residuals), then simple or ordinary kriging are applied to interpolate the residuals.
- Indicator Kriging, a non-parametric form of kriging in which the studied variable is first transformed in a binary variable (0,1) upon whether its values are above or below a given threshold. The resulting interpolation map would show the probabilities of exceeding (or being below) the threshold.

For an exhaustive overview of kriging and geostatistical analysis, please refer to:

Cressie, N., 1993. Statistics for Spatial Data, Wiley, New York

Deutsch, C. V., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and user's guide. Oxford University Press, New York.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Wackernagel, H., 2003. Multivariate Geostatistics. Springer, Berlin, Heidelberg.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists. Wiley & Sons Ltd., Chichester.

5.6.2.3 Regression and data mining techniques

In spatial statistics techniques (Figure 5.2) a statistical inference model between target variables and auxiliary variables (covariates) typically in feature space, is elaborated at the training locations, and then applied on a regular grid of unsampled points. The auxiliary variables are taken as representative of the factors of soil formation. The statistical inference model can be a regression model or a data mining (DM) model, either a decision tree, a machine learning, or a neural network type.

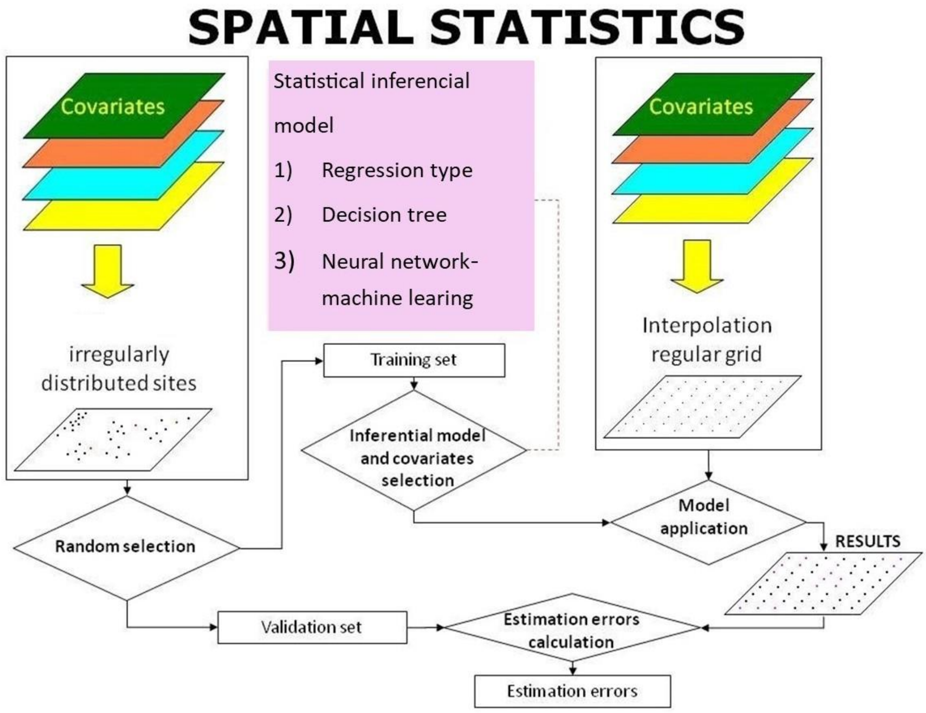


Figure 5.2 : Process of spatial statistics

DM algorithms are data-driven and originally developed for pattern recognition purposes. A benefit is that these algorithms can capture non-linear relationships and efficiently handle large datasets. As such they are typically well suited for deriving statistical relationships between soil properties and environmental covariates. However, as the algorithms are greedy for data, large soil databases are needed to build a robust prediction model. As such, the more conventional regression techniques and geostatistical methods remain vital in the toolkit of the digital soil mapper who has relatively few soil information available. For these reasons, both the conventional regression techniques and machine learning (ML) algorithms are discussed below.

In order to choose the best inference model, an exploratory performance comparison based on the available soil data and covariates can be performed on a validation dataset, to select the method giving the smallest validation errors (chapter 5.9). The generalization performance of any learner, including of course deep neural networks, is quantified by the difference between the training error and the test error. Good learners are those where the test error and the training error have close values (Theodoridis, 2020). DM usually perform better at the stage of training, and tend to overfit. When a model overfits, it learns to correctly predict all the training targets but fails when predicting validation or test targets. In other words, it fails to apply the knowledge gained during the training phase to data it has never seen before (i.e., it fails to generalize). We can spot overfitting by monitoring the training and validation errors. As soon as the validation error starts increasing (or remaining constant), whilst the training error continues decreasing, we can safely conclude that the model is overfitting. To mitigate this issue, we have three main options (or a combination thereof):

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

(i) utilize a larger data set (again, either by collecting more data or by artificially augmenting the available ones), (ii) lower the model complexity, or (iii) use a regularization term in the loss function (e.g. early stopping).

Regression techniques

The goal of regression is to predict the value of one or more target variables t given the value of a D-dimensional vector x of input variables. The polynomial is a specific example of a broad class of functions called linear regression models, which share the property of being linear functions of the adjustable parameters. The simplest form of linear regression models are also linear functions of the input variables. However, we can obtain a much more useful class of functions by taking linear combinations of a fixed set of nonlinear functions of the input variables, known as basis functions. Such models are linear functions of the parameters, which gives them simple analytical properties, and yet can be nonlinear with respect to the input variables. (Bishop et al., 2006)

Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. Linear regression is the most common regression analysis, in which the line - or a more complex linear combination - that most closely fits the data according to a specific mathematical criterion is found. In regression kriging, often the fixed trend is determined using a multiple linear regression model see e.g., the work of Samuel-Rosa et al., 2015. Alternatively, generalized linear models and linear mixed effect models have been used (Zang et al., 2020).

For further information please see the chapter on Regression in An Introduction to Statistical Learning[197].

Decision trees

Decision trees, including regression trees for numeric variables and classification trees for soil classes, have gained widespread use in DSM. Advantages of decision trees include the fact that they can use both numeric and categorical covariates, are computationally inexpensive, are somewhat robust towards redundant and/or correlated covariates and can handle nonlinear relationships and interactions between variables. Decision trees work by repeatedly splitting the dataset based on the covariates that give largest reduction in entropy (or other metric like Gini) in the resulting subsets. Single decision trees are highly prone to overfitting (when pruning is disabled), but it is possible to obtain a more robust prediction by combining predictions from several trees. Boosting achieves this by creating a sequence of decision trees, where each tree focuses on the observations which had the largest errors in the previous trees. Meanwhile, bagging creates independent decision trees from bootstrap samples of the original dataset. Random Forest (RF) (Breiman, 2001) is a form of bagging wherein each split uses only a random subset of the covariates in the dataset. In recent years, RF has been the most frequently used ML algorithm in DSM (Padarian, Minasny, & McBratney, 2020). Quantile regression forests (QRF) use the resampling procedure of RF to generate prediction quantiles as a measure of prediction uncertainties. Vaysse & Lagacherie (2017) showed that QRF can accurately

---

[197] https://link.springer.com/chapter/10.1007/978-1-4614-7138-7_3

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

estimate the uncertainties associates with predictions in DSM. Chen et al. (2019b) showed that Survival Random Forest can be used to deal with censored soil data, such as for example soil depth.

For further information please see the chapter on Tree-based Methods in An Introduction to Statistical Learning.[198]

Neural Networks machine learning

A Neural Network (NN) is a learning machine. The procedure used to perform the learning process is called a learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective (Haykin, 2009). In the Figure 5.3(a) is reported a scheme of a neuron as published by Haykin (2009). The neuron of the Figure 5.3(a) can be represented also with an architectural graph as in the Figure 5.2(b). Two general kinds of NN exist: parametric and non-parametric.

Neural networks with more than three hidden layers are considered deep learning, a fast-growing trend for spatial data analysis. Convolution neural networks allow to include information from the covariates not only in the pixel where the data point (i.e. soil profile) is, but also in the immediate neighbourhood, providing potential insight in the short-range variability (Zhu et al., 2017).

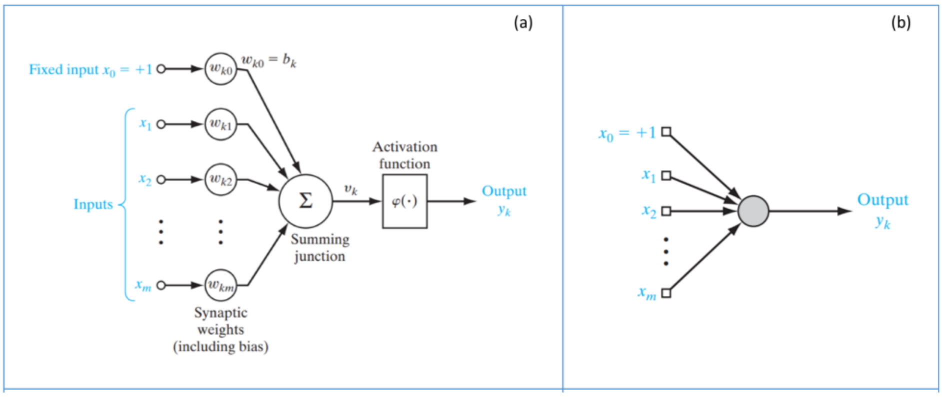


Figure 5.3 : Neural Network a,b (Haykin, 2009)

The parametric Neural Networks

The parametric NN are so-called because they are governed by a small number of adaptive parameters. The procedure is determining suitable values for the parameters, which leads to the minimizing of the error of estimation. Examples of such kind of Neural Networks are the Linear Neural Networks (LNN) and the Multilayer Perceptrons (MLP).

The LNN are Single-Layer Feedforward Networks (Figure 5.4), in which there is an input layer of source nodes that projects directly onto an output layer of neurons (computation nodes), but not vice versa.

198 https://link.springer.com/chapter/10.1007/978-1-4614-7138-7_8

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

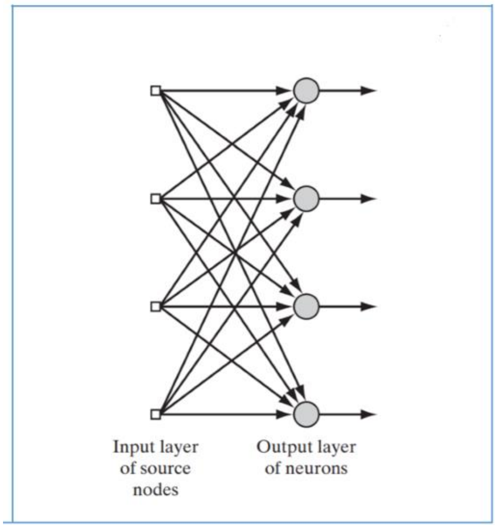


Figure 5.4 : Parametric Neural Network (Haykin, 2009)

The LNN are linear combinations of fixed nonlinear basis functions of the input variables (Bishop, 2006), which means that the activation functions are fixed before the training data set is observed. As a consequence, the number of basis functions needs to grow rapidly, often exponentially, with the dimensionality M of the input space.

The MLP networks are Multilayer Feedforward Networks, which are trained through an error backpropagation algorithm. The difference is that these functions are not fixed, instead they are trained by the system, and their parameters are adjusted through the error backpropagation algorithm. The architectural graph in Figure 5.5(c) illustrates the layout of a MLP network. The neurons in each layer of the network have as their inputs the output signals of the preceding layer. The Figure 5.5(d) illustrates the mechanism of forward propagation and error backpropagation.

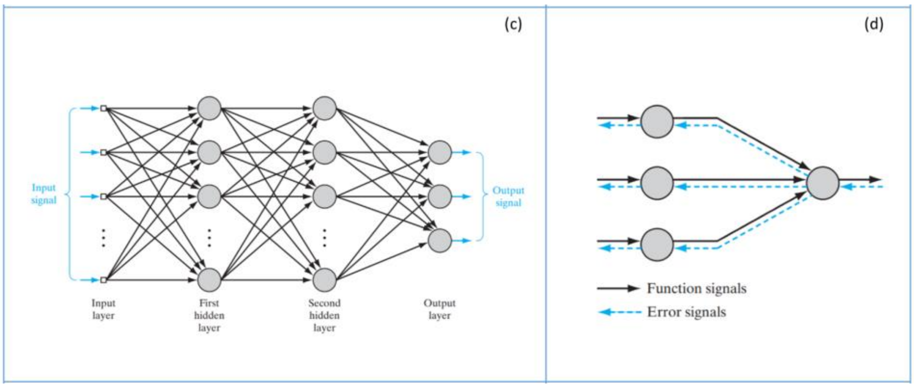


Figure 5.5 Neural Network c, d (Haykin, 2009)

The training algorithms of MLP network through error backpropagation involves an iterative procedure for minimization of an error function, with adjustments to the weights being made in a sequence of steps. We can distinguish between two distinct stages. In the first stage, the derivatives of the error function with respect to the weights are evaluated. In the second stage, the derivatives are then used to compute the adjustments to be made to the weights. The backpropagation learning algorithm for multilayer perceptrons may be viewed as the application of a recursive technique known in statistics as stochastic approximation.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement N° 862695

The non-parametric Neural Networks

These kinds of Neural Networks are called kernel methods, such as the Radial Basis Functions, or Support Vector Machines. The RBF networks were introduced for the purpose of exact function interpolation. Given a set of input vectors [x_1,x_2,…,x_N] along with corresponding target values d_1,d_2,…,d_N, the goal is to find a smooth function f(x) that pass through all the training data points and satisfies the interpolation condition F(x_i)=d_i for i = 1,2,...,N. This is achieved by expressing F(x) as a linear combination of radial basis functions, each one centered on every data point.

A RBF is constructed for every xi point, which in turn is taken as center.

The structure of the Radial Basis Function network consists of only three layers as depicted in the Figure 5.6.

a. *Input layer,* which consists of m_o source nodes, where m_o is the dimensionality of the input vector **x**.
b. *Hidden layer*, which consists of the same number of computation units as the size of the training sample, namely, N; each unit is mathematically described by a RBF. Unlike a MLP, the links connecting the source nodes to the hidden units are direct connections with no weights.
c. *Output layer*, which consists of a single computational unit.

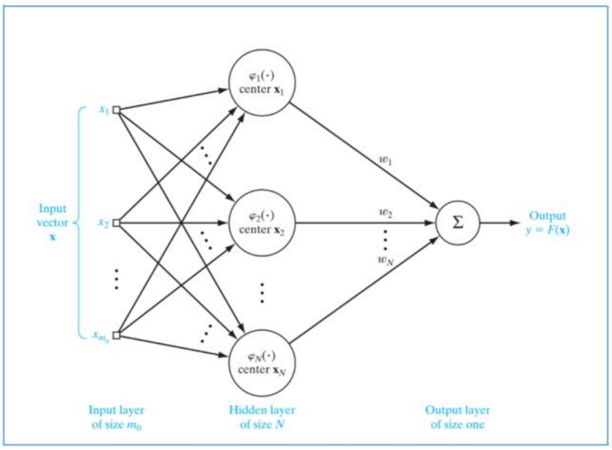


Figure 5.6 Radial Basis Function network (Haykin, 2009)

Other techniques for trend modelling

The concept of regression kriging (i.e. modelling trend and residuals separately) can be applied with many different techniques, often nonlinear. Most of the existing DSM studies not using a linear regression kriging or a tree-based approach are based on alternative methods such as Generalized Linear Models (e.g. Poggio and Gimona, 2017), Bayesian statistics (e.g. Poggio et al, 2019) and Bayesian Belief Networks (e.g. Taalab et al, 2015) or neural networks (e.g. Padarian et al, 2019, Wadoux, 2019).

5.6.3 Input data (how to pre-process)

Input data such as soil and soil forming factor data is represented in the regression matrix. The regression matrix is key to fitting a statistical model for DSM. The regression matrix contains information about the target variable and the corresponding explanatory variables. Using a clean and organized regression matrix, makes it is fairly easy to fit a statistical model between the target and explanatory variables in any programming language. In DSM, the target variable is typically a soil and the explanatory variables are the environmental covariates, as discussed in section 5.4. The regression matrix is tabular data, where there needs to be an entry for each individual soil sample and the corresponding covariate information. One can imagine that the regression matrix needs to be of good quality in order to avoid errors in the model and maps. Therefore, attention is needed to populate the regression matrix with clean and complete data.

5.6.3.1 In situ

Firstly, the regression matrix should contain information on the measured soil property. As a minimum, it must contain enough information and metadata (see chapter 3) to identify and locate each soil sample. Specifically, it should contain:

- The sample ID, so researchers can, if necessary, refer the observation to the soil database and other information associated with the sample, such as the sampling campaign, profile descriptions or laboratory methods.
- The sample depth, whether standardized or specific for each sample (so as on case of horizons). Often, the sample will represent a depth interval, and the regression matrix should include the depths of the upper and lower boundaries of this interval.
- The geographic coordinates of each observation. These are directly necessary for spatial interpolation and prediction. Furthermore, they are used for extracting values from the covariate layers to the regression matrix, and in some cases, they can be used as covariates on their own.
- The measured soil property. If the mapping effort aims to map several soil properties, it may be useful to include them all in the regression matrix at once.

Usually, DSM requires standardization of the sample intervals in order to map soil properties for specific depth intervals. One way to achieve this is by using weighted averages of the horizons that make up each of the standard depth intervals in a soil profile. Alternatively, it is possible to standardize depth intervals by using equal area splines to obtain a smooth function for the soil property with depth (Bishop et al., 1999, Malone et al., 2009). Some studies have bypassed depth standardization by using depth as a covariate (e.g., Ramcharan et al., 2018). However, Ma et al. (2021) cautioned against this approach, as they found that it produced “stepped” depth functions when combined with decision tree methods and had a lower accuracy than models based on standardized depths.

As many DSM efforts are based on legacy data collected over several campaigns, it can also be necessary to clean/harmonize (chapter 3) the data related to the observations, especially when it comes from different survey campaigns. This includes ensuring that the observations all refer to the same coordinate system and removing observations with unrealistic values. Additionally, some

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

observations may be located outside the areas covered by the covariate layers. In these cases, it may be necessary to remove the observations with missing covariate data or use imputation techniques to estimate missing data. Alternatively, if a covariate layer covers only a limited area, which would lead to the exclusion of a large number of observations, researchers may instead opt to omit the covariate in question, if they deem that it is relatively unimportant.

Another general principle when preparing the dataset, is to have it properly balanced for qualitative (categorical) variables. This means that all classes should be equally represented in the dataset.

5.6.3.2 Covariates such as satellite imagery, DEM, etc. proximal sensing

Environmental covariates used for DSM, supposed to explain part of the physical and chemical process governing soil spatial variation, are described in section 5.4. Once such covariates have been identified, their quality should be assessed to ensure the best data available are being used. Resolution is one of the most important attributes to consider when selecting data. Many high-resolution data sources are currently available, but they can provide “too much information” and add undesirable noise and/or excess data storage and processing time to analysis and modelling. The scale of physical features or properties on the landscape and the extent to which these have been captured in the soil sample data should be considered when choosing the most appropriate resolution. The types of resolution - spatial, spectral, temporal, and radiometric - must be considered.

- Spatial resolution applies to all gridded data sources and equates to grid cell size. When considering spectral data derived from remote sensing sources, spectral resolution refers to the number of bands and the spectral band widths a sensor provides.
- Temporal resolution indicates the time of year and frequency of image acquisition. Seasonality or repetition of image acquisition over several years may be important, especially if several images are mosaicked together. Ideally, the images for a mosaic should be acquired on or near the same date to minimize differences in atmospheric and Earth surface conditions. If data meeting those criteria are not available, and data from different years are used, the data used should at least be from the same time of year.
- Radiometric resolution, a special case of semantic resolution related to the bit-depth of an image, is an important, though rarely considered, spectral sensor attribute. It refers to the number of grey levels the sensor can potentially differentiate.

Data rarely are in an immediately usable format. Several issues may occur with them, but most can be resolved during pre-processing. Thus, some degree of pre-processing is typically needed before the data can be incorporated into analysis or modelling.

At the beginning, it is important to ensure that all spatial data have the same spatial projection (geographic vs. projected datum, etc.) for ease of processing. There are many software programs that can be used to define the projection and re-project the data – some good open source software programs are R[199], QGIS[200] and SAGA GIS (Conrad et al., 2015). Also, understanding the units of the

[199] https://www.R-project.org/ R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
[200] http://www.qgis.org/ QGIS Geographic Information System. QGIS Association

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

data and how to interpret them is important. If units between data sources are not compatible, values may need to be converted. Usually, this information can be found in the metadata of the source data. Data ranges should be noted, too.

When terrain attributes are needed, they can be derived from a DEM. DEMs are typically represented using the raster data format. Elevation can also be represented as points (e.g., LiDAR returns) or triangulated irregular networks (TIN), but the raster format is preferred due to its greater flexibility. Terrain attributes may be broadly grouped into two categories: (1) primary attributes, which are computed directly from a DEM; and (2) compound attributes, which are combinations of primary attributes (e.g. Compound Topographic Index, Moore et al., 1991). The field of geomorphometry (Hengl and Reuter, 2008) has advanced with the technology of GIS and is contributing to the evolving list of terrain attributes. Elevation data are developed to model the bare earth terrain features from a number of sources, and each source has a unique set of issues.

In case spectral imagery are used for DSM, several data quality checks and preprocessing steps are required. For spectral imagery, issues include clouds, smoke, sun glint, data loss, but also the calibration atmospheric and topographic correction algorithms used by the data provider. Images without clouds and smoke are preferable since these issues cannot be resolved through pre-processing. If an alternate image is not available, data pre-processing techniques should try to reduce the impact of sun glint, data loss, or calibration issues on analysis.

In Digital Soil Mapping, spectral data commonly is not used as raw spectral bands, but transformed to obtain useful band ratios related to specific characteristics of the earth surface, such as NDVI (Loiseau et al., 2019).

In short, some basic guidelines for data pre-processing should always be considered:

1. Ensure that all data are in the same projection and have the same extent.
2. Normalize spatial resolution (grid cell size) between layers.
3. If multiple datasets are being combined, it may be best if they share a common spatial resolution.
4. For elevation data, include in DEM preparation: filling sinks and trimming peaks, removing linear, human-made artefacts, applying a low-pass filter or other smoothing algorithm.
5. For spectral data, apply image standardization or atmospheric correction to calculate surface reflectance. If a mosaic is required, apply all the pre-processing prior to mosaicking.

More detailed information can be found in Hengl and Reuter 2008, the USDA site[201] and in Wadoux et al 2020.

5.6.4 How to choose a suitable method for mapping

This paragraph can be used as a quick guide to the essential steps or choices that should be taken to select the most appropriate or suitable procedure for mapping. It builds on the information provided in the other paragraphs and chapters of this report.

[201] https://www.nrcs.usda.gov/wps/portal/nrcs/detail/?cid=nrcs142p2_054255

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

1. Discuss with end-users: What is the purpose of the map, the expected quality, the accuracy needed. Decide for a CSM or DSM approach.
    a. The purpose of the map should be discussed with end-users to define what needs to be mapped (i.e, the target variables, soil properties or soil types), what is the expected or required quality, and if a map of accuracy is needed. For DSM, the resolution should be decided on and map scale in case of CSM.
    b. A reason to choose for CSM could be the desire to stratify the area in rather homogeneous spatial entities and understand their spatial distribution. End users can prefer a CSM map because they need to know how the soil types, and especially the *DerivedSoilProfiles*, are distributed in the soil-landscape space. For example, for a local soil service that is working with farmers, a map of soil types with their derived soil properties may be preferable to a DSM based soil properties maps because each soil type with its own *DerivedSoilProfile* has a specific preferred soil management system. DSM is suitable when enough data is available and especially for making continuous soil property maps. For soil type maps both approaches can be used, for DSM indicator kriging, disjunctive kriging, compositional kriging and more recent work on model-based geostatistics is available.
2. Exploratory analysis for DSM
    a. Which soil input data are available and is it suitable for use in the DSM analysis? This can depend on the spatial coverage, spatial correlation or sampling density? For judging the suitability of the available point and covariate data, the list by de Gruijter et al., (2006) can be used, described in chapter 4.2.
    b. Which potential covariates are available, are they suitable for DSM analysis and how do they relate to the soil data (SCORPAN model)?
    c. Is it possible to perform a new soil survey to improve the soil data available? If so, a sampling strategy can be chosen as described in chapter 4.
3. Formulate what you can deliver based on exploratory analysis
    a. If you have no spatial correlation in the soil data nor covariates that relate to the soil data than you can only do a deterministic interpolation or use a simple geostatistical method like ordinary kriging with a pure nugget semivariogram (see chapter 5.6.2.1).
    b. If the soil data set is small and doesn’t have any spatial correlation, then probably a conventional regression technique suffices but likely has low accuracy. If the model residual has no spatial correlation you cannot do much more (see chapter 5.6.2.3).
    c. If the soil data has spatial correlation and no correlation with available covariates, then ordinary kriging is most suitable (see chapter 5.6.2.2). If you know the mean, simple kriging will do.
    d. If soil data is correlated with covariates and the model residuals have spatial correlation, regression kriging or kriging with external drift is an option (see chapter 5.6.2.2).
    e. If you have a large dataset (both soil data and covariate data) then a data mining method, such as decision trees, or neural networks is an option. If there is still spatial correlation in model residuals than additional kriging of residuals might be worthwhile (see chapter 5.6.2.3). Another option for large data sets is to apply

special kriging algorithms like fixed rank kriging (Cressie & Wikle, 2011) or multi-resolution kriging (Lacoste et al., 2016).

A general workflow for DSM analysis has been proposed by Minasny and McBratney (2010) and used by the GlobalSoilmap project (Arrouays et al., 2014). An interactive online application for selecting the appropriate approach to spatial interpolation problems is the mapmakersguide.org[202].

5.6.5 Uncertainty maps

Practitioners often confuse resolution with accuracy. A high-resolution map is then erroneously interpreted as being more accurate than a low-resolution map. We are usually not only interested in the predicted spatial distribution of a soil property, but also in its accuracy. A map produced by block kriging (chapter 5.6.2.2) for instance has a higher accuracy than a map produced by punctual kriging given the same data and semivariogram. Intuitively that makes sense. Indeed, it is more ambitious to predict the value of a soil property at a specific point than to predict the average of an entire area. Aggregation leads to higher and disaggregation to lower accuracy.

Uncertainty maps of a specific map product are extremely useful. Uncertainty maps help us to judge the fitness-of-use of the map predictions for a specific question. They also reveal where additional sampling effort is required, that is, in areas where map quality is too low. Only an uncertainty map or estimate can inform on that. Higher resolution can mask higher uncertainty.

In CSM uncertainty maps can be obtained by producing *DerivedSoilProfiles*, with their associated standard deviations and 90% confidence intervals for quantitative properties (paragraph 5.5.2) Another indication for CSM maps is the map purity, if provided. Geostatistical interpolation methods not only *minimise* the variance of the prediction error, but also *quantify* this variance (e.g., Isaaks & Srivastava, 1986; Cressie, 1993; Goovaerts, 1997). Each location on the map therefore not only has a prediction, but also a measure of associated uncertainty. This can be the variance of the prediction error, *i.e.*, the usual by-product of kriging. In case of stochastic simulation, it can even be the entire spatially correlated conditional cumulative distribution function (ccdf, Goovaerts, 1997; Deutsch & Journel, 1998). The expectation of the ccdf is usually taken as prediction (*e.g.*, as in ordinary kriging). However, other measures, like the (conditional) median or mode are also possible. As a matter of fact, the ccdf makes it possible to estimate any quantile of the soil property. The ccdf can also be used to express our uncertainty about the prediction. Measures of spread like the variance, the standard deviation, the interquartile range (IQR), the 90% prediction interval width, or other differences between quantiles are usually used for this.

The uncertainty for a prediction can be estimated in spatial statistics methods using for instance bootstrapping, for example by applying the predUncertain function in soilassessment package in R to extract the uncertainty by a bootstrap approach (Efron, 1992). For some machine learning approaches also uncertainty information on the predictions is possible, for instance for quantile regression forests. Stochastic simulation makes it possible to generate numerous realizations of a soil property as a stack of maps. This stack resembles our uncertainty about the real value of the soil property. We may use map stacks like these as inputs to (non-linear) process models to quantify the effect of uncertainty in their inputs on their outputs (a technique known as error propagation). See also Deutsch & Journel

[202] https://www.mapmakersguide.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

(1998), Goovaerts, 1997. Uncertainty maps are also useful to perform risk-analysis. Goovaerts (1997) gives examples in which he uses uncertainty maps in combination with loss functions to minimize remediation costs of heavy metal pollution. Szatmári and Pásztor (2019) tested a procedure to compare various uncertainty modelling approaches by computing the actual fraction of true values falling inside the respective confidence intervals obtained with the different approaches.

A topic that deserves more attention is the presentation of uncertainty on a map. An uncertainty map is often presented next to the map with the predictions. End-users often only consider the map with the predictions and tend to ignore the uncertainty map. We should therefore consider additional ways to present the predictions and associated uncertainty preferably on a single map, or as multiple maps such as showing maps of the 0.05 and 0.95 quantile next to the prediction map, with the same colour legend (Arrouays et al., 2014; Heuvelink, 2014).

Typically, large-scale assessments rely on relatively sparse samples which makes the development of reliable and effective soil prediction models difficult. Grunwald et al. (2011) emphasized that the assessment of spatial and temporal autocorrelations depends on the density and distribution of soil observations within a landscape. Accordingly, the sample size of target variables and the sampled variability of the soil and environmental observations influence the accuracy of soil prediction models (Vasques et al., 2012). It is expected that datasets, where the density and scale of soil observations more closely resemble the spatial resolution of the SCORPAN factors, can elucidate the scaling behaviour of soil properties and processes across spatial and temporal scales (Mulder, 2013).

5.7 Sources of uncertainty

In digital soil mapping uncertainty arises from several soil data and mapping sources. Hengl and MacMillan (2019) describe possible sources of error for soil profile data as sampling (human) bias or omission of important areas, positioning error (location accuracy), sampling error (at horizon level), measurement error (in the laboratory), temporal sampling error, data input error (or typing error) and data interpretation error. If these types of error sources are expressed in a more compact way, it is possible to divide them into 4 main categories; uncertainty in the property measured in a soil measurement, e.g. field sampling, storage, lab measurement (i), positional uncertainty of soil measurement (ii), uncertainty of covariates (iii) and uncertainty in models (iv). The first group commonly contains errors in field and laboratory measurements. For instance, wrong sampling depth and procedures, ineliminable errors of field estimates because of difficulty in soil properties estimation or subjective estimates arising from soil scientist in the field. In the laboratory, analysis errors can occur due to differences in applied methods, interpretation of methods and quality of the analysis (chapter 3.4). Soil data collection is generally performed using geographic coordinates of the sampling or measurement locations. This is not a big problem when done with a good positional accuracy, using a high-precision GPS. But this is not always the case and therefore positional uncertainty may still influence a digital soil mapping model and may lead to a decline in the quality of the final soil map, especially when making high resolution maps. The positional accuracy of legacy soil data, sampled when GPS instruments were not available, is very low. On the other hand, soil legacy data can still be relevant for DSM because it adds to the data density and may help to cover the spatial variability of soil properties and covariate DSM model as well. Covariates maps used for modelling (chapter 5.4)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

such as a digital elevation model or land cover map could have errors and uncertainties in different degrees. The error of these maps can propagate and lead to larger uncertainties through the properties associated with it.

Model uncertainty refers to the model itself which also introduces uncertainties even if the soil data and covariates are without error. This is because modelling can only make an estimate of the real situation, it cannot be represented one-to-one. Although DSM models are suggested to represent soil mechanisms though extrapolation, in practice the creation of a model that entirely represents the real situation is impossible (Yigini et al.,2018).

5.7.1 Natural heterogeneity

Natural heterogeneity is not a criterion that is specific for conventional soil mapping (CSM) uncertainty, it is a general challenge for spatial soil analysis. In CSM the soil properties or class within a mapping unit vary within certain limits, well defined by a soil expert as map purity related to the chosen legend. This already indicates that soil properties may deviate from the specifications or legend made. The variation is strongly influenced by the landform, geological genesis, geological parent material and soil formation and also by the mapping resolution, which dictates the minimum size of the map unit or polygon. The soil or geo(morpho)logical factors are quite specific for different countries and their regions. Therefore, the required density of samples in the field for delineating soil mapping units cannot be recommended in a general sense. The tolerated variation of soil properties within a soil mapping unit is commonly focusing on agricultural practices or soil class maps (Knotters and van Egmond., 2019).

5.7.2 Scale and resolution

In CSM the tolerable variance of soil properties per area determines the resolution of field assessment and, consequently, the range of soil properties within a mapping unit. The required density of field measurements does not depend on the aimed resolution of the map. Small scale maps, for example, show in general a relatively low resolution due to clarity and comprehensibility. When it comes to downscaling, it is necessary that the resolution of the basic maps processed by disaggregation fulfil the requirements for the soil properties that are mapped.

5.7.3 Long term changes in soil conditions

As conventional soil mapping procedures on national scale took a long time or are still ongoing in some countries, major changes in the soil properties are most likely due to human activities. Land management practices, like terrain levelling, large scale draining measures - in Austria especially between the 1960ies and 1970ies, and soil erosion had a massive impact on soil water budget and on soil organic carbon losses in agricultural soils. Sealing of soils due to land-use changes became a severe issue in terms of qualitative soil protection. When it comes to compare uncertainties in conventional soil maps, the “time stamp” of the mapping is therefore of great importance.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

5.7.4 Methods – An example from Austria

Systematic and nationwide soil assessment and soil mapping were carried out in many countries in the second half of the past century. In Austria, for example, the mapping of agricultural soils on national level aiming at a resolution of 1:25 000 was started in the late 1950ies and is now being completed. Today, the main focus is to apply digital soil mapping techniques, to make the information available on the internet by using WEB GIS applications[203] and to improve the quality of conventionally assessed data by digital soil mapping (DSM) techniques. Sufficient manpower in the last decades allowed large scale assessment at a high resolution. Lack of modern survey equipment, like GPS, was not the main problem for uncertainties. Higher uncertainties were caused by transforming and downscaling polygons of field maps into the background map without using digital technologies. As the final analogue map was released at a predefined scale, the level of detail decreased; mapping units not reaching a minimum extent could not be represented, generalizations and aggregations had to be made. Unfortunately, the far more detailed and accurate field maps were usually not used as a basis for later digitization for cost reasons. To some extend also changes in analytical methods (e.g., for soil organic carbon) may also play a role.

5.7.5 CSM for evaluating grid-based soil information: An Austrian case study

The Austrian digital soil map was used to evaluate the representativeness of the European LUCAS project for Austrian circumstances. 80 LUCAS sample plots were selected, classified and analysed (soil chemistry, soil physics). It turned out, that in some region's samples in an area of 12,5 square meters around the LUCAS grid point showed already significant variances of key parameters, especially soil organic carbon. Within a circle of 100m radius around a LUCAS grid point up to 4 different soil units (even different soil types) in the digital soil map were observed. The preliminary results of this still ongoing study indicate that DSM techniques, well suited for improving and supporting conventional soil maps, cannot substitute lacking field data without regional soil expert knowledge.

5.8 Presence and causes of the transboundary problems

Various inconsistencies and discrepancies can exist in data or maps between countries, regions, or even surveys within particular countries (for sampling differences see chapter 4.3). They may be caused by several reasons (Borůvka et al., 2018), such as described in the following paragraphs.

5.8.1 Sampling scheme and density and the localization of samples

Sampling schemes and density can significantly differ between various countries and surveys. Though the sampling locations in national surveys are often distributed more or less evenly over the country, the particular locations are influenced by local conditions, so that the final distribution is nor regular, neither random. As a result, the proportion of various soil classes, land-use categories, or landscape units in the databases does not necessarily correspond to the real proportion of these classes. The

[203] https://bodenkarte.at/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

hardly defined sampling schemes may cause problems when using the data by current computation methods of DSM (chapter 4.2.2). For calibration, some methods for sampling from the database can be used to get a more representative dataset (e.g. Taghizadeh-Mehrjardi et al., 2020). The data are not well suitable for validation purposes either however, as the sampling is not random.

Exact localization of the pits or sampling locations is a problem especially in case of older surveys. Current surveys use precise localization by GPS, however, in the older surveys no such methods were available. Moreover, the coordinates are sometimes also jittered or truncated in new surveys to respect the European privacy legislation (General Data Protection Regulation - GDPR). By this, we can lose many of the data, or at least their exploitability is limited. Several ways have been used to determine or estimate the coordinates of the sampling locations as precisely as possible. First, the coordinates can be read from original maps where the sampling locations are indicated. However, the original marks in maps were done mostly by hand and might not be as precise as necessary. Second, the localization can be set-up according to the soil description in combination with available soil and land-use maps. The final position thus does not need to correspond to the real sampling location, but the conditions of the estimated position will correspond to the soil profile. The third method is the most costly and time consuming. It consists of searching the position in the field, where the current soil profile corresponds closely to that in the database. This method is used only exceptionally, usually in the frame of some regional projects aiming in the analysis of temporal development of soil properties, when new samples are collected at the same time for comparison with the old ones. Some computational methods have been also developed that can cope with such imprecise location of samples, examples are shown in Barber et al. (2006) or Cressie and Kornak (2003). Exact localization of the sampling locations is also complicated in case of composite samples, particularly in those cases when the support area for the composite sample is bigger. In some cases, data are assigned to polygons rather than points. One way to cope with this is by applying geostatistical disaggregation techniques such as area-to-point kriging and area-to-area kriging (Kyriakidis, 2004; Brus et al., 2014).

5.8.2 Sampling depth

A number of studies have been conducted to compare soil profile sampling according to horizons and sampling in fixed depths. Both have their advantages and disadvantages. The recalculation between the different depths is possible by mass-based weighted means, preferably using bulk density, but bulk density of all the depths should be known for this. In the GSM specifications, a depth function - particularly the spline function (Odgers et al. 2012) - is recommended for this purpose. Though it is not an ideal solution, it is probably the best one among other possibilities. An advantage is that in this case it does not matter whether the data come from fixed depths or individual horizons.

An even bigger problem is that many surveys focus only on the surface layers, topsoils only (e.g. LUCAS), or topsoil and subsoil, while the deeper layers are hardly sampled.

5.8.3 Extent of analyses and analytical methods

Each soil survey has a different set of analyses performed according to the objectives of the survey. While the basic soil properties like soil pH, soil organic carbon content, or available nutrients are often measured, some other data, especially the more demanding analyses (for example soil texture, sorption characteristics, content of specific pollutants, soil physical characteristics like bulk density or available water capacity etc.), are often missing.

Even if the required soil characteristics were determined, the methods of their determination are not unified and can differ between the countries and surveys, or could even change in one monitoring during the time (see chapter 3.5 for more information about harmonisation of lab methods). For example, soil organic carbon can be determined oxidimetrically by Walkley-Black method or modified Tyurin method, by loss on ignition, or by spectral methods. Various extractants are used for available nutrients (Mehlich 1, 2, 3, Egner, Schachtschabel, Bray, Olsen). Even exchangeable pH is determined differently, in KCl and $CaCl_2$ extracts of various concentrations. There are suggested recalculations in the GSM specifications (Science Committee 2015); however, they need to be verified on the local soils. Moreover, an additional uncertainty is introduced by their recalculation to the database again.

5.8.4 Aging of data

The problem of data aging is difficult to solve in many older legacy data. Not only some locations may be lost due to land-use change, but the values on identical locations can change. In temporally stable soil properties like soil texture the values do not change in time so much. It is more difficult in case of less stable properties and properties strongly influenced by soil management like soil pH, soil organic matter or available nutrient contents. One solution is field verification, when selected original locations are resampled, analysed and the results are compared to the original values (retrospective monitoring). If the changes are in some way consistent, they can be used for an update of data on other similar locations. However, this procedure is quite costly and time consuming, so that it could be applied only on a small part of the locations. The resulting new data can be used for trend analysis or be the beginning of a monitoring series (chapter 4.2.3.2).

5.8.5 Example of trans-boundary GSOC map border effect between Austria and Slovakia

Possible reasons of disagreement between national contributions of Austria (AT) and Slovakia (SK) to Global Soil Organic Carbon Map by FAO/GSP (Figure 5.7) were explored in 2018 as a part of a non-funded cross-border project under coordination of ESP Pillar 4 (chapter 1.2.9).

a) **b)** **c)**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

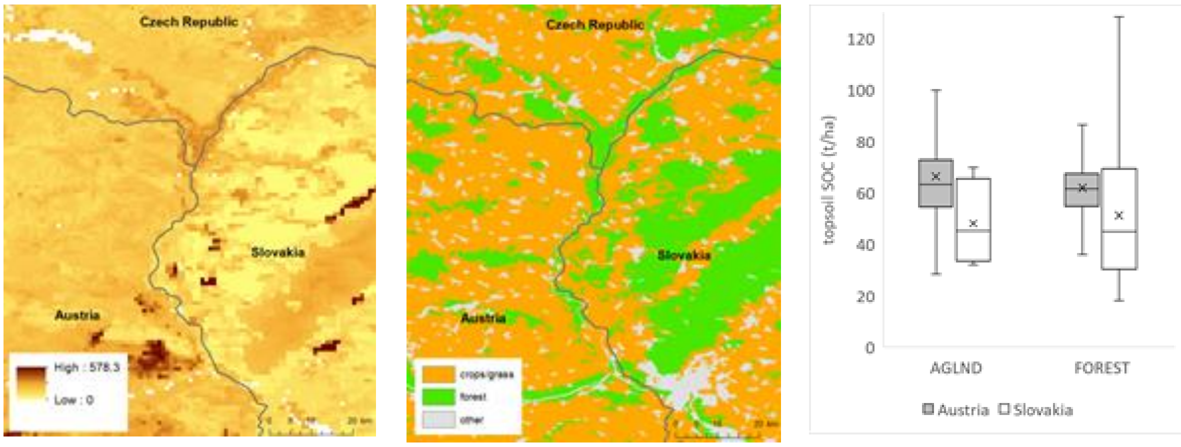


Figure 5.7 : Border effect on the boundary of AT and SK well visible on the GSOC 1.5.0 global map product as a result of map compilation from the national contributions: topsoil (0 – 30 cm) SOC stock in t/ha (a), simplified land cover classification of the cross-border region (b), and a distribution of reported topsoil SOC stock values separately, for agricultural land and forests in AT and SK (c).

The AT submission to the GSOC map was a result of a national SOC mapping project consisting of contributions from several key partners covering both agricultural and forestry sectors. Two major institutes for agricultural and forestry soil inventories and soil information administration contributed to the national SOC stock map of SK.

Meta-analysis of the input data and algorithms used for compiling the national contributions for GSOC map revealed that both in AT and SK methods used for SOC map compilation differed based on the focus area.

Class-area soil map quantification algorithm employing measured soil profiles as a source of soil analytical information (SOC concentration, soil texture) was used for estimating SOC stock in agricultural land, mostly relying upon dense coverage of soil observations and corresponding maps in detailed scale.

A limited number of soil observations from national soil monitoring programs in AT and SK were taken together with better spatially distributed data on landscape variables (covariates), and SOC stocks for forest soils were then estimated with use of geostatistical methods.

Possible reasons of trans-boundary inconsistency in estimated topsoil SOC values might reside in data sampling methods and sampling densities (but this in case of AT and SK was more pronounced for the national estimates for agricultural soils and forests), analytical methods and possible conversions of measured SOC concentration values in the input data processing (wet oxidation methods were used for earlier surveys, which were later replaced by dry combustion – with conversion assumed in AT, but not in SK), and possibly also soil classification and map compilation might have contributed to the differences in topsoil stocks estimated for AT and SK (with putting less emphasis on soil polygons and more relying on dense network of soil observations in SK as compared to AT approach preferring compilation of typical estimated soil profiles for mapped soil typological units).

Another harmonisation problem identified from the input data meta-analysis, which also might have partly influenced the level of consistency across (but also within) the national borders, was the temporal validity of input topsoil SOC data. All topsoil SOC data for agricultural land in SK was collected and analysed between 1961 and 1970, while in AT the interval for the agricultural soil data collection

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

was much wider, starting in early 1960s and ending up in 2000s, depending on the region. Forest soil data in SK was collected after 1990 and represent much more up-to-date information compared to agricultural land.

5.8.6 Transboundary problems in conventional landscape-based soil mapping

The production of traditional soil maps and databases at a national level poses a potential problem in the use of this data for international purposes. Nevertheless, the demand for detailed and reliable legacy data has increased substantially in the last decades, mainly through costly and time-consuming acquisition of primary soil data (Arrouays et al., 2017). National legacy data most often originates from traditional soil surveys and mapping campaigns, producing point soil profile observations and soil polygon maps. Generally, there are several general problems that need to be addressed when creating transboundary map products: i) different initial purposes and motivation (aims) for the soil survey. Soil survey and mapping campaigns have been motivated by various objectives, i.e., general description of soil resources, determining suitable land use, increase in agricultural production or soil protection policy. The objective for the soil survey affects the primary setting and design of the campaign (chapter 4), including the mapping scale, sampling density and the choice of surveyed soil attributes; ii) different detail of soil survey and mapping resulting in map products available in different scales and level of detail. This represents a major challenge in border areas where the map polygons need to be smoothly connected; iii) differences in settings and organization of mapping campaign. The conventional soil survey and mapping depend largely on the surveyor and his/her intuition, knowledge and ability to synthetize available information on the soil variability related to terrain, geology, land use etc. (Heuvelink and Webster, 2001). This expert knowledge enables the surveyor to select a most suitable "typical" site for the profile description. However, even this personal judgement can be objectified by the primary organization and formal requirements of the survey. These include the pre-set sampling design (general density and type of sampling, density of sampling in homogeneous and heterogeneous soil areas, refinement of polygon boundaries by additional sampling), availability of auxiliary data (geology, geomorphology maps, aerial photographs) and also the choice and training of the surveyors; iv) differences in soil profile description and samples processing resulting from different national standards applied in surveying, mapping and analysing of soil. These include a large set of possible discrepancies, such as actual soil profile description (designation of soil horizons, description of soil features such as soil colour or texture), soil classification system or laboratory methods. All these issues largely correspond to the period of the data acquisition and practical purpose of the survey.

The potential of the use of the data derived from regional soil survey largely depends on the possibility, extent and quality of the harmonization of the data with currently used soil description and classification systems. In the last decades, national soil databases have been harmonised to integrate into international databases (Arrouays et al., 2017; Waltner et al., 2012), develop global maps and models (Hessel et al., 2014), or investigate new possibilities of correlation among different classification systems (Hughes et al., 2017, Láng et al., 2013; Zádorová et al. 2021). The data harmonization methods frequently include reclassification of texture or skeleton classes, processing of colour to RGB schemes, horizon- to depth-structured profile organisation conversion using different methods of cluster analysis and depth functions, and transformation of analytical data (most often pH, cation exchange capacity, and soil organic carbon) measured by different methods (Beaudette et al., 2016; Hughes et al., 2017). Transformation of polygon-based maps of soil classes into an internationally usable format has been explored via various quantitative methods including taxonomic distance and inter-taxa variation (Minasny and McBratney, 2007; Láng et al. 2013; Hughes et al. 2017; Zádorová et al. 2020, 2021).

5.8.7 DSM techniques to deal with transboundary problems and inconsistencies

Various DSM tools have been developed and used for harmonization of the existing maps across borders, like rasterization and disaggregation (see chapters 5.3.2). In case of missing analyses, a solution is using some pedotransfer functions and rules. However, their calibration on local samples is necessary and, moreover, additional uncertainty is introduced to the data by this approach. An example is provided in paragraph 5.8.5.

5.9 DSM validation

5.9.1 Importance of validation

A digital soil property map is a spatial representation of the actual variation in soil properties, but like other representations in the form of maps or models, it is always a generalisation of reality. Map validation is therefore necessary to determine whether a digital soil map is accurate enough for a practical application, or to enable comparison of information accuracy between maps. Validation is generally done by comparing predicted values (e.g. soil property or soil class) against observed values at known positions. Comparisons at the calibration sites, known as internal accuracy, often results in overly optimistic results and the validation should therefore be made by external or test accuracy, using data not included in the calibration (Brus et al., 2011).

5.9.2 How to validate

Validation can be, and is, performed in many different ways. These can be divided into three main methods; i) probability sampling, ii) cross validation (CV), and iii) data splitting. The different methods have been summarized, tested and discussed in e.g. Muller et al. (2004), Brus et al. (2011), Schmidt et al. (2014), Biswas and Zhang (2018), and Piikki et al. (2021) and a summary is presented here.

5.9.2.1 Validation by probability sampling

Validation by probability sampling is the preferred validation method (Brus et al., 2011). Indeed, validation by means of purposive sampling may be problematic as has been demonstrated by Knotters and Brus (2013). One important advantage of probability sampling is the possibility to produce objective, unbiased and valid validation metrics. Sampling can be done at the same time as the calibration sampling but should preferably have its own design (chapter 4.2.1). Probability sampling not only results in a validation metric, but also in its standard error. As a result, the statistical significance of the validation metric can be tested.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

5.9.2.2 Cross validation

If time and money do not allow for an independent validation data set, leave-one-out cross validation or Bootstrapping is the second-best choice. Cross validation can be done in several ways. In leave-one-out CV (LOOCV) all samples are left out from the calibration and used later on for validation one at a time. In K-fold CV the data is divided into folds, or segments, with *n* samples in each. One segment is left out at the time and each sample appears in a validation fold once. Repeated k-fold CV is k-fold CV repeated *m* times (with differently composed folds) with each sample appearing in a validation fold once every *m*. LOOCV is slightly biased, but has high variability. K-fold CV has less variability but introduces bias (Wehrens et al., 2000). One typically uses K=5 or K=10, depending on the size of the dataset. Note that cross-validation as described above should not be confused with *spatial* cross-validation which is warned against by Wadoux et al., 2021a. Bootstrapping uses random sampling for validation from the total sample set with replacement repeated a large number of times. Because of the random sampling with replacement, a sample can appear in multiple validation sets, but also, though less likely, in none.

5.9.2.3 Data-splitting

One-time data splitting, if not random, can result in biased subsets and is therefore a less good choice. In studies with few samples, splitting the data into a validation and calibration set also leads to models based on an even smaller number of observations which can lead to less accurate and less robust models. The data can be split in a number of ways including systematic, random, random stratified or targeted in covariate space and random stratified in geography. If the original sample design is a probability sampling, using a random split into calibration and validation samples enables unbiased estimation of accuracy.

In a recent review of validation strategies for DSM of continuous variables the proportion of validation samples in the studies that used data-splitting varied largely, from 15% to 77% (Piikki et al., 2021a). The robustness of both calibration and validation is dependent on the number of samples used and how representative they are. It is therefore difficult to give a general recommendation of a proportion of calibration and validation samples. However, with few total samples available, having a larger proportion of the samples in the calibration set to ensure as representative observations as possible in the calibration may be preferable. However, for small data sets, methods like LOOCV are generally preferred.

5.9.2.4 Clustered samples and targeted sampling strategies

Extra care needs to be taken with spatially clustered data or data with multiple samples from the same profile. That is in cases where samples may be regarded as semi replicates of each other. In these cases, LOOCV and random split of the data, either ones or in k-fold CV, may lead to dependent validation sets and overestimated map accuracy (Lagacherie et al. 2020). Here LOOCV and random split is not recommended. Instead the data split in the k-fold CV should systematically leave one cluster or profile out at a time.

Strategies with targeted sampling are very common in DSM (). This may cause problems when the data is split into calibration and validation sets as the calibration data no longer will capture the designed variation and as a consequence validation statistics may possibly be misleading.

5.9.2.5 Sample support and scale issues

Spatial support for observed values (chapter 4.4.2) is important for interpretation of validation metrics (Malone et al., 2013; Bishop et al., 2015; Piikki and Söderström, 2019). If, for example, the validation data consist of laboratory analyses of soil samples with point support, the validation measure will show how well the map represents reality at point locations. If instead the validation samples represent one-hectare averages, the map accuracy for one-hectare areas is evaluated. Using point support may underestimate map accuracy at the relevant support for the intended map use, if the intended support is larger (Piikki et al., 2021). It is therefore relevant to indicate the used support with the validation metrics.

The validation should also be done at the same scale as the intended use. That means that if a national DSM is expected to be able to resolve within-field variation, validation needs to be done at the field scale. Map extent is also important to consider. For example, an R^2 for a continent does not tell how well the variation within a farm or a county is explained.

5.9.3 Validation metrics - aspects of error

In this section, we make a distinction between validation metrics for continuous variables and validation methods for categorical variables. For an overview of metrics being used for DSM and their basic features, we refer to Piikki et al. (2021) and the and in the R package ‘valmetrics’ (Piikki et al. 2021b). For an integrated approach for evaluating quantitative soil maps, we refer to Wadoux et al., (2021b).

5.9.3.1 Validation for continuous variables

Most validation metrics are functions of the predicted ($\hat{z}_i$) and observed values (z_i) at locations *i*.

The error (Ellili Bargaoui et al., 2019; Kato, 2016; Pal, 2017) at location *i* is given by:

$$e_i = \hat{z}_i - z_i$$

The mean error, systematic error or bias is given by the arithmetical average of e_i. Its range is $\mathbb{R}$, and its optimum is zero. A value greater than zero means overestimation and a value smaller than zero underestimation. Play attention because some authors use to calculate the mean error with the opposite formulation, that is observed minus predicted. In that case an opposite interpretation must be given. Decreasing the absolute value of the systematic error means increasing the trueness.

The variance of e_i is a measure of random error. For convenience, it is often expressed by its square root (*i.e.*, the standard deviation of e_i) as this quantity is in the same units of measurement as *z*. Its range is $\mathbb{R}_{\geq 0}$, with optimum value 0. Decreasing the random error means increasing the precision.

The mean squared error (MSE), *i.e.*, the arithmetical average of e_i^2 is a measure of accuracy. The MSE is the sum of the variance of e_i and the squared bias. For convenience, it is often expressed by its square root (*i.e.*, the root mean squared error, RMSE) as this quantity is in the same units of measurement as *z*. Its range is $\mathbb{R}_{\geq 0}$, with an optimum value of 0. The smaller the RMSE, the more accurate the map. To suppress the effect of outliers, the mean absolute error (MAE), *i.e.*, the arithmetical average of $|e_i|$, is sometimes used as an alternative (Isaaks & Srivastava, 1989). Figure 5.8 gives four scatter plots in which the observations are on the ordinate (y-axis) and the predictions (mapped values) on the abscissa (x-axis). The scatter plots in the top row are more precise than those in the bottom row. The scatter plots in the left column are more biased than those in the right column. The scatter plot on the upper right is the most accurate, the one in the lower left the most inaccurate.

Another interesting metric is the ratio of the variances of the predicted and observed values. This ratio indicates the degree of smoothing. Its range is $\mathbb{R}_{\geq 0}$. The optimal value is one (no smoothing, *i.e.*, the variation on the map is identical to the variation in the field). Values lower than one indicate smoothing, values greater than one indicate that the variation has been exaggerated on the map.

Pearsons (product moment) correlation coefficient between predicted and observed values can be used to validate spatial patterns on the map. It ranges from -1 to 1. The optimal value is 1. In that case spatial patters are fully reproduced by the map. A value of zero indicates absence of linear correlation. A value of -1 indicates that predicted and observed values have a negative linear relationship (are in anti-phase). Note that Pearsons correlation coefficient is insensitive to bias. A severely biased map may still have a Pearsons correlation coefficient near one. Therefore, Pearsons correlation coefficient should always be considered in conjunction with a metric that quantifies bias or accuracy. See also Wadoux et al. (2021b) for examples on validation of maps by comparing multiple metrics simultaneously.

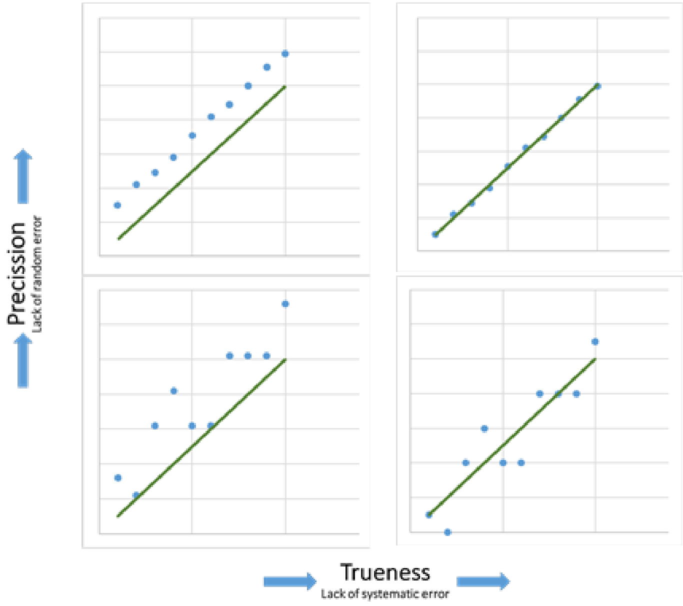


Figure 5.8 The concepts of precision, trueness, and accuracy explained by means of four scatter plots. See main text for details (5.9.3).

Some metrics are dependent on spread in data, number of samples, or unit. This is worth to consider in the choice of metrics as it has implications for their suitability to comparing the accuracy for mapped properties between each other or between models.

5.9.3.2 Validation for categorical variables

The validation of categorical variables is done with indexes based on confusion matrices. Given a categorical variable with n classes (x=1…n), a confusion matrix is constructed, which is a two-dimensional matrix, with in one dimension the observed classes and in the other dimension the predicted classes (Ting K.M., 2011). Usually it is constructed such that each row represents the instances in the predicted class, and each column represents the instances in the observed class.

The instances that are correctly and wrongly predicted are named as (Bethanney et al., 2020):

1) the true predicted (TP) instances are the number of cases correctly predicted for a class x.

2) the false predicted (FP) instances are the number of cases wrongly predicted for a class x.

3) the false non-predicted (FN) instances are the number of cases wrongly non predicted for a class x.

4) total predicted (P) instances are the total number of cases predicted for a class x.

5) total observed (O) instances are the total number of cases observed for a class x.

6) total non-observed (N) instances are the total number of cases non observed for a class x, that is all the cases minus the total number of cases observed for a certain class.

7) prevalence (prev) is the ratio between the number of observed cases belonging to the specific class and the total number of cases.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

		Observed			
		x=1	*x=...*	*x = n*	*P*
Predicted	*x=1*	$\boldsymbol{TP_{x=1}}$	$\boldsymbol{FP_{x=1}}$ $\boldsymbol{FN_{x=\cdots}}$	$\boldsymbol{FP_{x=1}}$ $\boldsymbol{FN_{x=n}}$	$\boldsymbol{P_{x=1}}$
	x=...	$\boldsymbol{FP_{x=\cdots}}$ $\boldsymbol{FN_{x=1}}$	$\boldsymbol{TP_{x=\cdots}}$	$\boldsymbol{FP_{x=\cdots}}$ $\boldsymbol{FN_{x=n}}$	$\boldsymbol{P_{x=\cdots}}$
	x = n	$\boldsymbol{FP_{x=n}}$ $\boldsymbol{FN_{x=1}}$	$\boldsymbol{FP_{x=n}}$ $\boldsymbol{FN_{x=\cdots}}$	$\boldsymbol{TP_{x=n}}$	$\boldsymbol{P_{x=n}}$
	O	$\boldsymbol{O_{x=1}}$	$\boldsymbol{O_{x=\cdots}}$	$\boldsymbol{O_{x=n}}$	*All cases*

The following indexes can be derived from the confusion matrix, all ranging from 0 to 1, with 1 being the most correct and 0 the most incorrect.

The producer accuracy or sensitivity (Se) of the model for each class x, which is calculated as the ratio between the number of cases correctly predicted in a specific class and the total number of observed cases belonging to that class.

$$Se = TP_x/O_x$$

The user accuracy (Ua) of the model for each class x, which is calculated as the ratio between the number of cases correctly predicted in a specific class and the total number of cases predicted by the model for that class.

$$Ua = TP_x/P_x$$

The specificity (Sp) of the model for each class x, which is calculated as the ratio between the number of cases correctly predicted by the model as not belonging to a specific class and the total number of observed cases not belonging to that class.

$$Sp = (All\ cases - TP_x - FP_x - FN_x\ )/All\ cases - O_x$$

The overall accuracy (Oa) of the model, which is calculated as the ratio between the total number of cases correctly predicted and the total number of cases.

$$Oa = \sum_{x=1}^{n} TP_x/All\ cases$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

The positive (pred+) and negative (pred-) predictivity of the model for each class x, which is calculated with the Bayesian formulation (Lesaffre and Lawson, 2012) as follow.

$$pred+ = \frac{Se * prev}{Se * prev + (1 - Sp) * (1 - prev)}$$

$$pred- = \frac{Sp * (1 - prev)}{Sp * (1 - prev) + (1 - Se) * prev}$$

More recently, alternatives for the confusion matrix have been proposed based on taxonomic distances (see Minasny et al., 2010; Rossiter et al., 2017).

5.10 Grid specifications

To facilitate making comparable maps and a subsequent compilation of maps made by different countries into a joint product common grid specifications have been defined within Europe and globally. The specifications describe for instance the resolution, scale, projection, units, etc. Some national specifications additionally require a specified minimum accuracy or map purity. This requires an adequate calibration and validation exercise (Knotters 2019). European and global specifications do not, they only indicate resolution. Important to note is that a high or low resolution of a map does not in any way indicate its accuracy. Unless an uncertainty map or measure is provided it is not possible to assess its accuracy, for instance by looking at the resolution.

In Europe these are the INSPIRE grid specifications (paragraph 5.10.1). Globally, there are two initiatives (paragraph 5.2) that defined specifications for global maps (paragraph 5.10.2). The first is the GlobalSoilMap initiative and the second are the specifications adopted by the Global Soil Partnership for its Global Products such as the GSOC map and others. In recent years (2017-2019) these two initiatives have started a collaboration where GlobalSoilMap will advise in the grid specifications for the Global Soil Partnership.

Within EJP SOIL we will follow the INSPIRE specifications.

5.10.1 INSPIRE

The INSPIRE specification for grids are given in the INSPIRE Technical Documentation "D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines" version 3.1[204] :

"Either of the grids with fixed and unambiguously defined locations defined in Sections 2.2.1 and 2.2.2 shall be used (IR Requirement, Annex II, Section 2.2, Grids) as a geo-referencing framework to make gridded data available in INSPIRE, unless one of the following conditions holds:

[204] https://inspire.ec.europa.eu/id/document/tg/gg

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

(1) Other grids may be specified for specific spatial data themes in Annexes II-IV. In this case, data exchanged using such a theme-specific grid shall use standards in which the grid definition is either included with the data, or linked by reference.
(2) For grid referencing in regions outside of continental Europe Member States may define their own grid based on a geodetic coordinate reference system compliant with ITRS and a Lambert Azimuthal Equal Area projection, following the same principles as laid down for the grid specified in Section 2.2.1. In this case, an identifier for the coordinate reference system shall be created.

Section 2.2.1, Equal Area Grid

The grid is based on the ETRS89 Lambert Azimuthal Equal Area (ETRS89-LAEA) coordinate

reference system with the centre of the projection at the point 52o N, 10o E and false easting: x0 = 4321000 m, false northing: y0 = 3210000 m. The origin of the grid coincides with the false origin of the ETRS89-LAEA coordinate reference system (x=0, y=0).

Grid points of grids based on ETRS89-LAEA shall coincide with grid points of the grid. The grid is hierarchical, with resolutions of 1m, 10m, 100m, 1000m, 10000m and 100000m. The grid orientation is south-north, west-east. The grid is designated as Grid_ETRS89-LAEA. For identification of an individual resolution level the cell size in metres is appended.

Section 2.2.2, Zoned Geographic Grid

(1) When gridded data is delivered using geodetic coordinates the multi-resolution grid defined in this Section may be used as a geo-referencing framework. (…)
(2) The grid shall be based on the ETRS89-GRS80 geodetic coordinate reference system.
(3) The origin of the grid shall coincide with the intersection point of the Equator with the Greenwich Meridian (GRS80 latitude ϕ=0; GRS80 longitude λ=0).
(4) The grid orientation shall be south-north and west-east according to the net defined by the meridians and parallels of the GRS80 ellipsoid.
(5) For grid referencing in regions outside of continental Europe data providers may define their own grid based on a geodetic coordinate reference system compliant with ITRS, following the same principles as laid down for the Pan-European Grid_ETRS89-GRS80zn. In this case, an identifier for the coordinate reference system and the corresponding identifier for the grid shall be created."

5.10.2 GSM/GSP for other products.

For GSP global products, the generic grid at 30 arc-seconds resolution (approximately 1 x 1 km) has been provided by ISRIC World Soil Information (Yigini et al., 2018).

In GlobalSoilMap, a 3 arc -second grid is adopted (approximately 100 x 100 m), which is defined to exactly match the NASA Shuttle Radar Digital Elevation Model grid (extended north and south to the poles) (GlobalSoilMap, 2015).

5.11 Choices within EJP SOIL

In this chapter we have tried to provide an overview of mapping approaches, choices to be made, possible data to use, sources of uncertainty and how to calculate and address them, and ways to validate the maps. Different intended uses for a map and the availability of existing (in situ point or covariate) data will result in different preferred approaches, there is not one best method. To make a soil class map at field or local scale or a soil property map at national or European scale will most likely incite very different approaches. At the same time there are a few general best practices that we advise to adhere to.

We encourage to follow the process elaborated in chapter 5.6.4 and start with defining the purpose of the map and inventorying the existing data. Thereby using knowledge of soil forming factors and the SCORPAN model to make effective choices and verifying the quality of the input data and eliminating possible errors, i.e. 'garbage in is garbage out'. During the entire data collection and mapping process a good documentation of methods, metadata, sampling design strategy and protocol, used data, chosen method, resulting uncertainty metrics and maps, validation of the result and a continuous effort to decrease possible sources of uncertainty are very important and result in a better quality map and repeatable mapping process. This includes using knowledge and experience from projects, initiatives as described in chapter 1, together with literature that is cited throughout the entire report (chapter 6). This also includes using existing data as inventoried in chapter 2, a good data organisation as described in chapter 3 and a sampling strategy as described in chapter 4, informed by the most suitable mapping method as elaborated in chapter 5.

Within EJP SOIL we will adhere to the INSPIRE grid specifications for European and national maps and will aim to reduce transboundary inconsistencies and/ or propose possibilities to address these such as (lab) method harmonisation, combining different sample designs and protocols, using GPS etc.

Within EJP SOIL we aim to choose the methods for the entire soil information workflow wisely depending on the aim, use the latest insights. We aim to document carefully and carry out uncertainty analyses and map validation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

6 References

6.1 Chapter 1 - references

AlZu'bi, S., Hawashin, B., Mujahed, Y., Jararweh, B., and Gupta, B.B., 2019. An Efficient Employment of Internet of Multimedia Things in Smart and Future Agriculture. Multimedia Tools and Applications, Vol. 78, October 1, 2019.

Augier, J. , Papuchon, J., Devot, A., Ayral, V., De Pierrepont, A., Marechal, A., Meredith, S., Lisbona, D., Poirier, E., Maire, E., Pinsault, D., Frelih-Larsen, A., Clement, T., 2020. Evaluation study on the impact of the CAP on the sustainable management of soil. Final Report to DG Agriculture. Alliance Environnement.

Baritz, R., Hudson, G., Willer J. and authors of the GS Soil test cases, 2012. Data Harmonization Best Practice Guidelines. D4.3 of the GS Soil project, funded under the eContentplus programme, OJ L 79, 24.3.2005.

Bas-Defossez, F. and Meredith, S., 2019. CAP 2021-27: Comparative Analysis of Environmental Performance of COMENVI and COMAGRI Reports. Report for NABU by the Institute for European Environmental Policy. 21 p.

Batjes, N.H., 2019. Technologically achievable soil organic carbon sequestration in world croplands and grasslands. Land Degradation & Development 30, 25-32. https://dx.doi.org/10.1002/ldr.3209

Batjes, N.H., Ribeiro, E. and Van Oostrum, A., 2020. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth Syst. Sci. Data 12, 299-320. https://doi.org/10.5194/essd-12-299-2020

CIRCASA, 2020. Deliverable D3.1: 'Strategic Research Agenda on soil organic carbon in agricultural soils.' European Union's Horizon 2020 research and innovation programme grant agreement No 774378 - Coordination of International Research Cooperation on soil CArbon Sequestration in Agriculture. https://doi.org/10.15454/LSWRDG

Dazzi, C. and Monteleone, S., 2007. Anthropogenic processes in the evolution of a soil chronosequence on marly-limestone substrata in an Italian Mediterranean environment. Geoderma 141(3–4), 201-209. https://doi.org/10.1016/j.geoderma.2007.05.016

Elbersen, B., Van Eupen, M., Mantel, S., Verzandvoort, S., Boogaard, H., Mucher, S., Cicarreli, T., Elbersen, W., Bai, Z., Iqbal, Y., Cossel, M., MCallum, I., Carrasco, J., Ciria Ramos, C., Monti, A., Scordia, D., Eleftheriadis, I., 2018. Deliverable 2.6: Methodological approaches to identify and map marginal land suitable for industrial crops in Europe. MAGIC; GA-No.: 727698. https://doi.org/10.5281/zenodo.3539311

European Commission, 2006. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - Thematic Strategy for Soil Protection [SEC(2006)620] [SEC(2006)1165] COM/2006/0231 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52006DC0231

European Commission, 2019. The European Green Deal. COM(2019) 640 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019DC0640&qid=1615372201861

European Commission, 2020. Future of the common agricultural policy. Briefs summarising the nine proposed specific objectives of the future CAP. https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/future-cap/key-policy-objectives-future-cap_en

European Commission, 2020a. A Farm to Fork Strategy for a fair, healthy and environmentally-friendly food system; Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and Committee of the Regions; COM(2020) 381 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0381

European Commission, 2020b. EU Biodiversity Strategy for 2030 - Bringing nature back into our lives; Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and Committee of the Regions; COM(2020) 380 final, 22 p. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0380&qid=1615372154418.

European Commission, 2020d. How the future CAP will contribute to the EU Green Deal. Factsheet, 2 p. Retrieved from: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/key_policies/documents/future-cap-and-green-deal_en.pdf

EEA (European Environmental Agency), 2020. The European environment —state and outlook 2020 Knowledge for transition to a sustainable Europe. Luxembourg: Publications Office of the European Union. https://doi.org/10.2800/96749.

GlobalSoilMap.net 2011. Specifications - Version 1 GlobalSoilMap.net products (Release 2.1), 50 p. http://www.globalsoilmap.net/system/files/GlobalSoilMap_net_specifications_v2_0_edited_draft_Sept_2011_RAM_V12.pdf

ISO 28258, 2013. Soil quality — Digital exchange of soil-related data. https://www.iso.org/standard/44595.html. Accessed on 10-02-2021.

Keesstra, S.D., Muro, M., Maring, L., Arellano Jaimerena, B., Van Eupen, M., Elbersen, B., McNeill, A., Tugran, T. and Markowska, A., 2020. Providing Support in Relation to the Implementation of Soil and Land Related Sustainable Development Goals at EU Level: Final Report, Wageningen Environmental Research Rapport: 3032, Wageningen Environmental Research.

Ma, Y.X., Minasny, B., Malone, B.P. and McBratney, A.B., 2019. Pedology and digital soil mapping (DSM). European Journal of Soil Science. March 2019, 70, 216 –235. https://onlinelibrary.wiley.com/doi/abs/10.1111/ejss.12790

Mission Board for Soil health and food, 2020. Caring for soil is caring for life – Ensure 75% of soils are healthy by 2030 for healthy food, people, nature and climate. Interim report for the European Commission, Directorate-General for Research and Innovation and Directorate-General for Agriculture and Rural Development. First edition, 56 pp. https://op.europa.eu/en/publication-detail/-/publication/4ebd2586-fc85-11ea-b44f-01aa75ed71a1/

Montanarella, L. and Panagos, P., 2021. The relevance of sustainable soil management within the European Green Deal. Land Use Policy, 100, http://dx.doi.org/10.1016/j.landusepol.2020.104950

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. and Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. European Journal of Soil Science 69, 140-153. http://dx.doi.org/10.1111/ejss.12499

Panagos, P., Jones, A., Van Liedekerke, M., Orgiazzi, A., Lugato, E. and Montanarella, L., 2020. JRC Support to the European Joint Programme for Soil (EJP SOIL) - Datasets, Technical Advice and Scientific Guidance. Technical Report by the Joint Research Centre (JRC), EUR 30450, Publications Office of the European Union, Luxembourg. https://esdac.jrc.ec.europa.eu/node/68798 (accessed 15-03-2021)

Poppe, K., and Vrolijk, H., 2016. Farm Sustainability Data for Better Policy Evaluation with FADN. Edited by K. Poppe and H. Vrolijk, The Hague, The Netherlands.

Rumpel, C., Amiraslani, F., Chenu, C., Garcia Cardenas, M., Kaonga, M., Koutika, L.-S., Ladha, J., Madari, B., Shirato, Y., Smith, P., Soudi, B., Soussana, J.-F., Whitehead, D. and Wollenberg, E., 2019. The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. Ambio. https://doi.org/10.1007/s13280-019-01165-2

De Sousa, L., Poggio, L., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E. and Rossiter, D., 2020. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL Discussion https://doi.org/10.5194/soil-2020-65

TWG-Soil, 2013. D2.8.III.3 Data Specification on Soil – Technical Guidelines, INSPIRE Thematic Working Group Soil, European Commission Joint Research Centre.

Wilkinson, M.D., 2016. Comment: The FAIR Guiding Principles for Scientific Data Management and Stewardship. Nature Publishing Group, 2016.

Yigini, Y. and Van Egmond, F., 2020. GLOSIS development. 6th Meeting of the International Network of Soil Information Institutions (INSII). GSP / FAO, Rome (7-9 October 2020). https://www.slideshare.net/Soils2012/item-i-ia-e-glosis-development-yusuf-yigini-fenny-van-egmond (accessed 15-03-2021)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

6.2 Chapter 2 - references

Bakacsi, Zs., Laborczi, A., Szatmári, G., Horel, Á., Dencső, M., Molnár, S., Ujj, E., Tóth, E., 2020. Compiling C/N and total-N dataset to support countrywide soil nutrient emission models for Hungary. Studies in Agricultural Economics 122, 86–95. https://doi.org/10.7896/j.2037.

Bieganowski, A., Ryżak, M., Sochan, A., Barna, G., Hernádi, H., Beczek, M., Polakowski, C., Makó, A., 2018. Laser diffractometry in the measurements of soil and sediment particle size distribution. Advances in Agronomy 151, 215–279. https://doi.org/10.1016/bs.agron.2018.04.003.

Lopatka, Artur., 2017. Europejski monitoring użytkowania gruntów i baza danych glebowych LUCAS. Studia i Raporty IUNG–PIB. ZESZYT 51(5): 73–89. https://doi.org/10.26114/sir.iung.2017.51.05.

McBratney, A., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Mulder V.L., M. Lacoste, A.C. Richer-de-Forges, D. Arrouays, 2016. GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. Science of the Total Environment, Vol. 573, 1352-1369. https://doi.org/10.1016/j.scitotenv.2016.07.066.

Panagos, P., Jones, A., Van Liedekerke, M., Orgiazzi, A., Lugato, E. and Montanarella, L., 2020. JRC support to the European Joint Programme for soil (EJP SOIL), EUR 30450 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-25322-8, doi:10.2760/74273, JRC122248.

Ramos, T.B., A. Horta, M.C. Gonçalves, F.P. Pires, D. Duffy, J.C. Martins, 2017. The INFOSOLO database as a first step towards the development of a soil information system in Portugal. CATENA, 158. 390-412. https://doi.org/10.1016/j.catena.2017.07.020.

Sagris, V., Devos, W., 2008. LPIS core conceptual model. Methodology for feature catalogue and application schema. JRC Scientific and Technical Reports. Publications Office of the European Union, Luxembourg. https://doi.org/10.2788/80091.

Tol-Leenders, D. van, M. Knotters, W. de Groot, P. Gerritsen, A. Reijneveld, F. van Egmond, H. Wosten, P. Kuikman, 2019. Koolstofvoorraad in de bodem van Nederland (1998-2018); CC-NL. Wageningen, Wageningen Environmental Research, Rapport 2974. https://doi.org/10.18174/509781

Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. Hydrological Processes 31, 2662–2666. https://doi.org/10.1002/hyp.11203.

Weynants, M., Montanarella, L., Tóth, G., Arnoldussen, A., Anaya-Romero, M., Bilas, G., Børresen, T., Cornelis, W., Daroussin, J., Feichtinger, F., Gonçalves, M., Hannam, J., Haugen, L.E., Hennings, V., Houškova, B., Iovino, M., Javaux, M., Keay, C., Kätterer, T., Kvaernø, S., Laktinova, T., Lamorski, K., Lilly, A., Makó, A., Matula, S., Morari, F., Nemes, A., Nyborg, Å., Patyka, N.V., Riley, H., Romano, N., Schindler, U., Shein, E., Slawinski, C., Strauss, P., Tóth, B., Wösten, H., 2013. European HYdropedological Data Inventory (EU-HYDI). Scientific and Technical Research series, EUR 26053 EN. Publications Office of the European Union, Luxembourg. https://doi.org/10.2788/5936.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

6.3 List of database references, cited by Partners

Austria

BAW specialized project data and long term research. https://www.baw.at/en/wasser-boden-en/projekte.html (accessed 26 February 2021).

BORIS Soil Information System. https://www.umweltbundesamt.at/boris (accessed 26 February 2021).

eBOD Digital Soil Map of Austria. https://bodenkarte.at (accessed 26 February 2021).

IfÖL–Long term research site. https://boku.ac.at/nas/ifoel/arbeitsgruppen/ag-bodenfruchtbarkeit-und-anbausysteme/abgeschlossene-projekte (accessed 26 February 2021).

Invekos–Agricultural Data. https://www.bmlrt.gv.at/land/direktzahlungen/Invekos.html (accessed 26 February 2021).

Österreichische Bodenschätzung (Austrian Soil Condition Survey). https://www.bmf.gv.at/themen/steuern/immobilien-grundstuecke/grundbesitzabgaben-einheitsbewertung/land-und-forstwirtschaftliches-vermoegen-bodenschaetzung.html (accessed 26 February 2021).

Belgium Flanders

Bodemafdekkingskaart 2015 (BAK), 5 m resolutie (Soil sealing rate). http://www.geopunt.be/catalogus/datasetfolder/9d3df29f-ac3b-497d-9ed7-097e37232151 (accessed 26 February 2021).

Bodemassociatiekaart (1:500.000). https://www.geopunt.be/catalogus/datasetfolder/c4f51b28-51bf-4189-8e98-72b94ae8da13 (accessed 26 February 2021).

DOV Soil Database for Flanders. https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/3f507fd9-24c0-40ab-9328-29f0dff571fe (accessed 26 February 2021).

Gekarteerde grondverschuivingen (Mapped landslides). https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/5e1146a5-4f72-4e93-a71d-8297d3dfec6c (accessed 26 February 2021).

Gevoeligheid voor grondverschuivingen (Susceptilbity landslides). https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/2035e01e-eef5-4806-948c-df7b966169a1 (accessed 26 February 2021).

Landbouwgebruikspercelen. http://www.geopunt.be/catalogus/datasetfolder/7cc9babc-e021-46bc-abb6-1b74d44b14ea (accessed 26 February 2021).

Potential Soil Erosion Map of Flanders. https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/298d34a4-d459-4fc8-a9be-e8e26cd9b354 (accessed 26 February 2021).

Soil Map of Flanders (1:20.000). https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/a1547a01-b9fc-40fa-a2eb-009a39c02c7b (accessed 26 February 2021).

Soil Organic Carbon Stock Maps for Belgium (40m grid). https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/bf8c35ad-ab45-4084-b1e1-3042c66204aa (accessed 26 February 2021).

Uitgevoerde gemeentelijke erosiebestrijdingswerken. https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/692a9efb-8509-4a6f-a3be-6887f969867d (accessed 26 February 2021).

WRB Reference Soil Groups 250k. https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/64e8a11d-5602-4ad5-b370-e4d396fe218f (accessed 26 February 2021).

WRB Soil Units 40k. https://www.dov.vlaanderen.be/geonetwork/srv/dut/catalog.search#/metadata/83c46eae-a202-454c-b063-a858be3e4335 (accessed 26 February 2021).

Belgium Wallonia

Chartin, C., Kruger, I., Van Wesemael, B., Carnol, M., 2017. Carbiosol, Carbone organique, biomasse et activité microbienne des sols: vers un indicateur de la qualité des sols en Wallonie. Rapport final. Belgium: Service public de Wallonie

Legrain, X., Demarcin, P., Colinet, G., Bock, L., 2011. Cartographie des sols en Belgique: aperçu historique et présentation des travaux actuels de valorisation et de révision de la Carte Numérique des Sols de Wallonie. Biotechnol. Agron. Soc. Environ. 15(S2), 647–656.

REQUASUD database. https://www.requasud.be/outils/ (accessed 26 February 2021).

Van Orshoven, S., Vandenbroecke, D., 1993. Guide de l'utilisation de Aardewerk – Base de données de profils pédologiques. Institut pour l'encouragement de la Recherche scientifique dans l'Industrie et l'Agriculture, Bruxelles

Czech Republic

Bonitované půdně ekologické jednotky (Evaluated soil ecological units). https://www.spucr.cz/bpej (accessed 26 February 2021).

Komplexní průzkum půd (Systematic soil survey). https://kpp.vumop.cz/?core=account (accessed 26 February 2021).

Land Registry LPIS (Land Parcel Identification System)–AZZP. http://eagri.cz/public/web/mze/farmar/LPIS/ (accessed 26 February 2021).

Poláková, S., Kubík,L., Prášková, l., Malý, S., Němec, P., Staňa, J., 2017. Monitoring Zemědělských Půd V České Republice 1992–2013 (Monitoring of agricultural soils in the Czech Republic 1992–2013) Ústřední kontrolní a zkušební ústav zemědělský (Central Institute for Supervising and Testing in

Agriculture) Brno http://eagri.cz/public/web/file/520923/Publikace_25_let_Monitoring_Zemedelskych_pud_1992_2013.pdf (accessed 26 February 2021).

Denmark

Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS ONE 9(8), e105519. https://doi.org/10.1371/journal.pone.0105519.

Danish Farmers' Registrations. https://lbst.dk/landbrug/kort-og-markblokke/markkort-og-markblokke/ (accessed 26 February 2021).

Greve, M.H., Christensen, O.F., Greve, M.B., Kheir, R.B. 2014. Change in peat coverage in danish cultivated soils during the past 35 years. Soil Science 179 (5), 250–257. http://dx.doi.org/10.1097/ss.0000000000000066.

Madsen, H. B., Jensen, N. H., 1985. The establishment of pedological soil databases in Denmark. Geografisk Tidsskrift 85, 1–8, Copenhagen. https://doi.org/10.1080/00167223.1985.10649211.

Madsen, H. B., Jensen, N. H., Jakobsen, B.H., Platou, S.W. 1985. A method for identification and mapping potentially acid sulfate soils in Jutland, Denmark. Catena, 12 (1), 363–371. https://doi.org/10.1016/S0341-8162(85)80031-X.

Madsen, H.B., Nørr, A.H., Holst, K.A., 1992.The Danish soil classification. The Royal Danish Geographical Society, Copenhagen, Denmark.

Taghizadeh-Toosi, A., Olesen, J.E., Kristensen, K., Elsgaard, L., Østergaard, H.S., Lægdsmand, M., Greve, M.H., Christensen, B.T., 2014. Changes in carbon stocks of Danish agricultural mineral soils between 1986 and 2009. The European Journal of Soil Science 65 (5), 730–740. https://doi.org/10.1111/ejss.12169.

Estonia

Large-scale Soil Map of Estonia. https://geoportaal.maaamet.ee/est/Ruumiandmed/Mullastiku-kaart-p33.html (accessed 26 February 2021).

Riiklik mullaseire keskkonnaseire andmete kogumise ja avalikustamise infosüsteemis KESE (National Soil monitoring in environmental monitoring data collection and disclosure information system KESE. https://www.keskkonnaagentuur.ee/et/eesmargid-tegevused/keskkonnaseire/otsi-seireandmeid (accessed 26 February 2021).

Soil erosion in Estonia modelled by use of USLE. https://pmk.agri.ee/et/pollumajanduskeskkonna-uuringud/uurimisvaldkonnad/muld (accessed 26 February 2021).

Finland

Palojärvi, A., Nuutinen, V. 2002. The soil quality concept and its importance in the study of finnish arable soils. Agricultural and Food Science 11(4), 329–342. https://doi.org/10.23986/afsci.5737

France

Carte de la Réserve Utile en eau issue de la Base de Données Géographique des Sols de France. https://doi.org/10.15454/JPB9RB (accessed 26 February 2021).

Carte du nom de sol dominant en WRB issue de la Base de Données Géographique des Sols de France. https://agroenvgeo.data.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/f599548d-b0be-55f9-9adf-4e9aa3a0be91 (accessed 26 February 2021).

Krigged maps of trace elements contents in French soil. https://doi.org/10.15454/VN9F6H. (accessed 26 February 2021).

Maps of the agronomic soil properties from BDAT. https://doi.org/10.15454/SVDTOU. (accessed 26 February 2021).

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016. GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. Science of The Total Environment 573, pp. 1352–1369. https://doi.org/10.1016/j.scitotenv.2016.07.066.

Réseau de Mesures de la Qualité des Sols (French Soil Monitoring Network). https://www.gissol.fr/le-gis/programmes/rmqs-34 (accessed 26 February 2021).

Román Dobarco, M., Bourennane, H., Arrouays, D., Saby, N.P.A., Cousin, I., Martin, M.P., 2019. Uncertainty assessment of GlobalSoilMap soil available water capacity products: A French case study. Geoderma 344, 14–30. https://doi.org/10.1016/j.geoderma.2019.02.036.

Soil geographical database of France at 1:1,000,000. https://doi.org/10.15454/BPN57S.(accessed 26 February 2021).

Soil test database. https://webapps.gissol.fr/geosol/ (accessed 26 February 2021).

Germany

German Agricultural Soil Inventory. https://www.thuenen.de/en/ak/projects/agricultural-soil-inventory-bze-lw/ (accessed 26 February 2021).

Landwirtschaftszählung/Agrarstrukturerhebung. https://www.destatis.de/EN/Themes/Economic-Sectors-Enterprises/Agriculture-Forestry-Fisheries/_node.html (accessed 26 February 2021).

Schneider, F., Don, A., 2019. Root-restricting layers in German agricultural soils. Part I: extent and cause. Plant and Soil, 442, 433–451. https://doi.org/10.1007/s11104-019-04185-9.

Soil map by the Federal Institute for Geosciences and Natural Resources (BGR) in cooperation with the National Geological Surveys (SGD) of the federal states and General Geological Map of the Federal Republic of Germany. https://geoviewer.bgr.de/mapapps4/resources/apps/geoviewer/index.html?lang=en (accessed 26 February 2021).

Various interpolated grids characterizing weather and climate in Germany. https://opendata.dwd.de/climate_environment/CDC/grids_germany/ (accessed 26 February 2021).

Hungary

Digital, Optimized, Soil Related Maps and Information in Hungary. http://dosoremi.hu/ (accessed 26 February 2021).

Farmer Diary Program on the Web (Webes Gazdálkodási Napló Program). https://gn.eurofarmer.hu/ (accessed 26 February 2021).

Hungarian Identification System of Agricultural Parcels. https://www.mepar.hu/ (accessed 26 February 2021).

Hungarian Soil Information and Monitoring System. http://terkep.air.gov.hu/website/nyilvanos_tim/viewer (accessed 26 February 2021).

Ireland

Irish Soil Information System. http://gis.teagasc.ie/soils/ (accessed 26 February 2021).

National Soil Database of Ireland. http://www.epa.ie/pubs/reports/research/land/nsdbfinalreport/ (accessed 26 February 2021).

Soil Quality Assessment Research Project. https://www.teagasc.ie/environment/soil/research/square/ (accessed 26 February 2021).

Tellus Geophyical and Geochemical Survey. https://www.gsi.ie/en-ie/programmes-and-projects/tellus/Pages/default.aspx (accessed 26 February 2021).

Italy

Costantini, E.A., Barbetti, R., Fantappiè, M., L'Abate, G., Lorenzetti, R., Magini, S., 2013. Pedodiversity, in: Costantini, E., Dazzi, C. (Eds), The Soils of Italy. World Soils Book Series. Springer, Dordrecht, pp. 105–178. https://doi.org/10.1007/978-94-007-5642-7_6.

Costantini, E.A.C., Fantappié, M., L'Abate, G. 2013. Climate and Pedoclimate of Italy, in: Costantini, E., Dazzi, C. (Eds), The Soils of Italy. World Soils Book Series. Springer, Dordrecht, pp. 19–37. https://doi.org/10.1007/978-94-007-5642-7_2.

CREA AA Italian historical wheater series. https://www.reterurale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/16319 (accessed 26 February 2021).

Rete di Informazione Contabile Agricola (Farms Agronomical Mangement database). https://gaia.crea.gov.it/#Rica (accessed 26 February 2021).

Latvia

Amelioration cadastre information system. https://www.melioracija.lv/ (accessed 26 February 2021).

Digital soil database (soil polygons). https://geolatvija.lv/geo/p/317 (accessed 26 February 2021).

Digital soil database (soil profiles). https://geolatvija.lv/geo/p/319 (accessed 26 February 2021).

Lithuania

Spatial Information Portal of Lithuania. https://www.geoportal.lt/geoportal/ (accessed 26 February 2021).

Netherlands

Basisregistratie Ondergrond - Borehole research - profile descriptions. https://basisregistratieondergrond.nl/inhoud-bro/registratieobjecten/bodem-grondonderzoek/booronderzoek-bhr (accessed 26 February 2021).

Basisregistratie Ondergrond - Soil face research. https://basisregistratieondergrond.nl/inhoud-bro/registratieobjecten/bodem-grondonderzoek/wandonderzoek-sfr/ (accessed 26 February 2021).

Basisregistratie Ondergrond - Soil map. https://basisregistratieondergrond.nl/inhoud-bro/registratieobjecten/modellen/bodemkaart-sgm/ (accessed 26 February 2021).

Bodemfysische eenheden kaart. https://www.wur.nl/nl/show/Bodemfysische-Eenhedenkaart-BOFEK2020.htm (accessed 26 February 2021).

Bongiorno, G., 2020. Novel soil quality indicators for the evaluation of agricultural management practices: a biological perspective Front. Agr. Sci. Eng. 7(3), 257–274. https://doi.org/10.15302/J-FASE-2020323.

Carbon Change dataset of the Netherlands. https://research.wur.nl/en/publications/koolstofvoorraad-in-de-bodem-van-nederland-1998-2018-cc-nl (accessed 26 February 2021).

Chris, J., Koopmans, C.J., Bloem, J., 2018. Soil quality effects of compost and manure in arable cropping. Results from using soil improvers for 17 years in the MAC trial. Louis Bolk Institute https://www.louisbolk.org/downloads/3331.pdf (accessed 26 February 2021).

Mol, G., Spijker, J.H., van Gaans, P., Römkens, P.F.A.M., 2012. Geochemische bodematlas van Nederland, Wageningen Academic Publishers, Wageningen. https://doi.org/10.3920/978-90-8686-743-1.

Publiek-Private Samenwerking Beter Bodembeheer – Bodemkwaliteit. https://www.beterbodembeheer.nl/upload_mm/4/1/c/35acd049-f895-45a2-9e60-d889b307a19e_Jaarverslagen%20Bodemkwaliteit%20Veenkolonien%202018%20Samengevoegd%202019.pdf (accessed 26 February 2021).

Norway

Avalanche and rockslide - rough susceptibility zones. https://kartkatalog.geonorge.no/metadata/avalanche-and-rockslide-rough-susceptiblitity-zones-wfs/18258934-0a8f-491b-9607-a3cb5754fa12 (accessed 26 February 2021).

Bedrock, scale 1:50,000. https://kartkatalog.geonorge.no/metadata/bedrock-scale-150000/a5c76d05-33bd-4a1d-b28b-81575092e468 (accessed 26 February 2021).

Digital terrain model 10x10 meters. https://hoydedata.no/LaserInnsyn/ (accessed 26 February 2021).

Flood hazard maps are prepared for floods with 20-, 200- and 1000-year return period. https://kartkatalog.geonorge.no/metadata/flood-hazard-map/e95008fc-0945-4d66-8bc9-e50ab3f50401 (accessed 26 February 2021).

Hydrology - mean annual runoff 1x1 km. https://kartkatalog.geonorge.no/metadata/hydrology-mean-annual-runoff/85a3486c-6d30-4cfb-957d-ae776e473ba1 (accessed 26 February 2021).

Incentive Programs related to land use and animal husbandry. https://data.norge.no/datasets/46850f03-98eb-4689-af4d-33594984cf2f (accessed 26 February 2021).

Land resource map 1:25.000. https://kartkatalog.geonorge.no/metadata/ar50-land-resource-map-150-000/4bc2d1e0-f693-4bf2-820d-c11830d849a3 (accessed 26 February 2021).

Land resource map 1:5.000. https://kartkatalog.geonorge.no/metadata/fkb-ar5-land-resource-map-15000/243751e8-5803-4627-898c-d0ddabe82056 (accessed 26 February 2021).

Norwegian soil survey sample grid 9 x9 km on agricultural land. https://kartkatalog.geonorge.no/metadata/soil/c961484d-9d4b-4f6a-8bd5-e145e96d1560 (accessed 26 February 2021).

Observations and climate data. http://www.senorge.no/aboutSeNorge.html (accessed 26 February 2021).

Superficial deposits 1:50 000 / 1:250.000; Marine limit (ML) indicates the highest level reached by sea after last glaciation 1:50 000. https://www.ngu.no/en/topic/datasets (accessed 26 February 2021).

Poland

Dobers, E.S., Ahl, C., Stuczyński, T., 2010. Comparison of Polish and German maps of agricultural soil quality using GIS. Journal of Plant Nutrition and Soil Science 173(2), 185–197. https://doi.org/10.1002/jpln.200800317.

Monit41k; MonitFarm. http://pw.iung.pl/index.php?option=com_content&view=article&id=182Itemid%3D99&Itemid=99 (accessed 26 February 2021).

Monitoring chemizmu gleb ornych Polski. http://www.gios.gov.pl/chemizm_gleb/ (accessed 26 February 2021).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Siebielec, G., 2017. Stały monitoring gleb użytków rolnych Polski. (Monitoring of the Chemical Properties of Arable Soils in Poland). Studia i Raporty IUNG-PIB 51(5) 57–72. http://www.iung.pl/sir/zeszyt51_4.pdf

Portugal

Database of Soil Profile Data for Portugal. https://data.isric.org/geonetwork/srv/por/catalog.search#/metadata/25d0cf4d-1865-4d2a-be32-40a1b2483936 (accessed 26 February 2021).

de Vos, B., Cools, N., 2011. Second european forest soil condition report. Volume I: Results of the BioSoil soil survey, INBO - Research Institute for Nature and Forest, Bruxelas. Report nº INBO.R.2011.35 https://pureportal.inbo.be/portal/nl/publications/second-european-forest-soil-condition-report(3732cc04-4a21-43b1-a2b5-2d68b18cf5da).html

Gonçalves, M. C., Ramos, T. B., Martins, J. C., 2013. Soil data from Portugal, in: Weynants et al. (Eds.), European Hydropedological Data Inventory (EU-HYDI). Publications Office of the European Union, Luxembourg. pp. 57–62. http://dx.doi.org/10.2788/5936.

Land Use and Soil Occupation Map for Portugal, 2018. https://snig.dgterritorio.gov.pt/rndg/srv/por/catalog.search#/map (accessed 26 February 2021)

Public irrigated areas and SIR (Irrigation Information System), 2018. https://sir.dgadr.gov.pt/expl-centro (accessed 26 February 2021).

Slovakia

Kobza, J., 1999. Legend for the 1:400 000 Soil Map of the Slovak Republic : A brief explanation of the constituent soil associations. VUPOP, Bratislava.

Kobza, J., Barančíková, G., Dodok, R., Makovníková, J., Pálka, B., Styk, J., Širáň, M., 2018. Land degradation and data reporting based on soil monitoring system in Slovakia. Innovative Techniques in Agriculture 2(5), 460–466. https://scientiaricerca.com/sritag/SRITAG-02-00063.php

Kobza, J., Barančíková, G., Makovníková, J., Pálka, B., Styk, J., Širáň, M., 2017. Current state and development of land degradation processes based on soil monitoring in Slovakia. Agriculture 63(2), 74–85. https://doi.org/10.1515/agri-2017-0007.

Makovníková, J., Širáň, M., Houšková, B., Pálka, B., Jones, A., 2017. Comparison of different models for predicting soil bulk density. Case study – Slovakian agricultural soils. International Agrophysics 31, 491–498. https://doi.org/10.1515/intag-2016-0079.

National soil monitoring system. https://envirozataze.enviroportal.sk/ (accessed 26 February 2021).

Skalský, R., Balkovič, J., 2002. Digital database of selected soil profiles of complex soil survey of Slovakia (KPP-DB), in: Proceedings No. 25. SSCRI, Bratislava, pp. 129–140.
Soil ecological units (land evaluation system). http://www.podnemapy.sk/portal/verejnost/bpej/bpej.aspx (accessed 26 February 2021).

Soil Map of Slovakiain 1:400k scale. http://www.podnemapy.sk/poda400/viewer.htm (accessed 26 February 2021).

Slovenia

Pedosequences of Slovenia. https://www.kis.si/en/Soil_data (accessed 26 February 2021).

Protected areas of Slovenia. https://www.gov.si/assets/ministrstva/MOP/Publikacije/4a9bb5ad59/zavarovana_obmocja.pdf (accessed 26 February 2021).

Soil Map of Slovenia. https://esdac.jrc.ec.europa.eu/projects/esoter/Danube/Presentations/c%20-%20Slovenia%20-%20Marko%20Zupan%20-%20SoilData_SLO_DanubeBasin_2015.pdf (accessed 26 February 2021).

Soil Pollution Monitoring. http://soil.bf.uni-lj.si/rots/index.php?page=gis/slo (accessed 26 February 2021).

Vrščaj, B., 2017. Soil Information, in: The Soils of Slovenia. Springer Netherlands, Dordrecht, pp. 157–170.

Vrščaj, B., Vernik, T., Bergant, J., Čuden, I., 2010. Vzpostavitev sistema multidisciplinarnih informacij prostora za napovedovanje in ocenjevanje škod po naravnih nesrečah v kmetijstvu. Delovni sklop 1, Izboljšava podatkov tal. Kmetijski inštitut Slovenije, Oddelek za kmetijsko ekologijo in naravne vire, Center za tla in okolje. (Multidisciplinary spatial infomation for prediction and assessment of damage after natural disasters in agriculture; droughts - soil water requrements. Agricultural Institute of Slovenia, Department for Agroecology and Natural Resources, Centre for Soil and Environment), Ljubljana, Slovenia.

Water protection areas of Slovenia. http://kazalci.arso.gov.si/en/content/water-protection-areas-0?tid=16 (accessed 26 February 2021).

Zorn, M., Komac, B., 2007. Probability modelling of landslide hazard. Acta Geogaphica Slovenica 47, 139–169. https://doi.org/10.3986/AGS47201.

Zupan, M., Rupreht, J., Tič, I., Šporar, M., Istenič, B., Lisec, A., 2006. Izdelava digitalne karte talnega števila. Biotehniška fakulteta Univerze v Ljubljani. Ljubljana.

Spain

CSIC-LTE CEBAS Tres Caminos. http://www.cebas.csic.es/ (accessed 26 February 2021).

CSIC-LTE ICA La Poveda. https://www.ica.csic.es/ (accessed 26 February 2021).

Guardia, G., Tellez-Rio, A., García-Marco, S., Martin-Lammerding, D., Tenorio, J.L., Ibáñez, M.Á., Vallejo, A., 2016. Effect of tillage and crop (cereal versus legume) on greenhouse gas emissions and Global Warming Potential in a non-irrigated Mediterranean field. Agriculture, Ecosystems and Environment 221, 187–197. https://doi.org/10.1016/j.agee.2016.01.047.

Martín-Lammerding, D., Navas, M., del Mar Albarrán, M., Tenorio, J.L., Lucena, J.J., 2015. Long term management systems under semiarid conditions: Influence on labile organic matter, β-glucosidase activity and microbial efficiency. Applied Soil Ecology 96, 296–305. https://doi.org/10.1016/j.apsoil.2015.08.021.

Martín-Lammerding, D., Tenorio, J.L., Albarrán, M.M., Zambrana, E., Walter I., 2013. Influence of tillage practices on soil biologically active organic matter content over a growing season under semiarid Mediterranean climate. Spanish Journal of Agricultural Research 11(1), 232–243. https://doi.org/10.5424/sjar/2013111-3455.

Martin-Rueda, I., Muñoz-Guerra, L.M., Yunta, F., Esteban, E., Tenorio, J.L, Lucena, J.J., 2007. Tillage and crop rotation effects on barley yield and soil nutrients on a Calciortidic Haploxeralf. Soil and Tillage Research 92(1-2), 1–9. https://doi.org/10.1016/j.still.2005.10.006.

Santín-Montanyá, M.I., Gandía-Toledano, M.L., Zambrana, E., Tenorio, J.L., 2020. Effects of tillage systems on wheat and weed water relationships over time when growing together, in semiarid conditions. Ann Appl Biol. 177, 256–265. https://doi.org/10.1111/aab.12620.

Santín-Montanyá, M.I., Martín-Lammerding, D., Walter, I., Zambrana, E., Tenorio, J.L, 2013. Effects of tillage, crop systems and fertilization on weed abundance and diversity in 4-year dry land winter wheat. European Journal of Agronomy 48, 43–49. https://doi.org/10.1016/j.eja.2013.02.006.

Tellez-Rio, A., Vallejo, A., García-Marco, S., Martin-Lammerding, D., Tenorio, J.L., Rees, R.M., Guardia, G., 2017. Conservation Agriculture practices reduce the global warming potential of rainfed low N input semi-arid agriculture. European Journal of Agronomy 84, 95–104. https://doi.org/10.1016/j.eja.2016.12.013.

Sweden

Djodjic, F., Markensten, H., 2019. From single fields to river basins: Identification of critical source areas for erosion and phosphorus losses at high resolution. Ambio, 48, 1129–1142. https://doi.org/10.1007/s13280-018-1134-8.

Geofysiska flygmätningar, gammastrålning (detaljerad) Airborn gamma radiation measurements. http://www.sgu.se/om-sgu/verksamhet/kartlaggning/geofysik_att_se_ner_i_berget/flyggeofysisk-matning/ (accessed 26 February 2021).

GSD-Elevation data, grid 2+.https://www.lantmateriet.se/sv/Kartor-och-geografisk-information/geodataprodukter/produktlista# (accessed 26 February 2021).

Jordlagerföljder (Soil layer data). http://resource.sgu.se/dokument/produkter/jordlagerfoljder-beskrivning.pdf (accessed 26 February 2021).

Meteorologisk analysmodell MESAN (AROME) – API. http://opendata.smhi.se/apidocs/metanalys/index.html (accessed 26 February 2021).

National Arable Soil & Crop Inventory. https://www.slu.se/institutioner/mark-miljo/miljoanalys/akermarksinventeringen/undersokningar/mark-grodoinventeringen/ (accessed 26 February 2021).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

National Inventory of Arable Soils. http://www2.jordbruksverket.se/download/18.4288f19214fb7ec78849af18/1441973777932/ra15_19.pdf (accessed 26 February 2021).

National Land Cover Database (NMD). https://gpt.vic-metria.nu/data/land/NMD/NMD2018_ProductDescription_ENG.pdf (accessed 26 February 2021).

National Soil Compaction Survey. https://www.slu.se/institutioner/mark-miljo/forskning/jordbearbetning/miljoovervakning/?si=C8E21EFA1402F7CC038D7AAE7BF5EF7F&rid=2023297005&sn=sluEPi6-prodSearchIndex (accessed 26 February 2021).

Observationsfält på jordbruksmark (Monitoring water from arable fields); Typområden på jordbruksmark (Monitoring water from agricultural catchments). https://www.slu.se/institutioner/mark-miljo/forskning/hydroteknik/observationsfalt-typomraden/ (accessed 26 February 2021).

Piikki, K., Söderström, M., 2019. Digital soil mapping of arable land in Sweden - Validation of performance at multiple scales. Geoderma, 352, 342–350. https://doi.org/10.1016/j.geoderma.2017.10.049.

Switzerland

Swiss Soil Dataset Version 5, 2020. https://nabodat.ch/index.php/de/service/bodendatensatz (accessed 26 February 2021).

Turkey

Boron Map of Turkey's Soils. https://arastirma.tarimorman.gov.tr/toprakgubre/Haber/155/Enstitumuz-Tarafindan-Ilk-Olarak-Uretilmis-Olan-Turkiye-Topraklari-Bor-Haritasi-Guncellenmistir (accessed 26 February 2021).

National Soil Carbon Stock Map; Salinity Map of Turkey's Soils; P_2O_5 Map of Turkey's Soils. https://arastirma.tarimorman.gov.tr/toprakgubre/Belgeler/Karbon_Panel/Toprak%20Organik%20Karbon%20Bilgi%20Sistemi%20Projesi.pdf (accessed 26 February 2021).

Turkish Ministry of Agriculture and Forestry Soil Fertilizer and Water Resources Central Research Institute Soil Information System. http://85.25.185.76/tgskmae/starter.aspx (accessed 26 February 2021).

United Kingdom

AFBI Pesticide usage monitoring surveys. https://www.afbini.gov.uk/articles/pesticide-usage-monitoring-surveys (accessed 26 February 2021).

AFBI Soil Series Map of Northern Ireland; AFBI 5k Soil Attribute Dataset. https://www.afbini.gov.uk/articles/soil-maps-and-soil-survey (accessed 26 February 2021).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

CEH Land Cover Map 2017. https://www.ceh.ac.uk/services/land-cover-map-2007 (accessed 26 February 2021).

CEH Land Cover plus Crops. https://www.ceh.ac.uk/crops2015 (accessed 26 February 2021).

CEH Land Cover plus Fertilisers; CEH Land Cover plus Pesticides. https://www.ceh.ac.uk/services/ceh-land-cover-plus-products-fertilisers-pesticides (accessed 26 February 2021).

Countryside survey of topsoil in Great Britain. https://catalogue.ceh.ac.uk/eidc/documents?facet=topic%7C0%2FSoil%2F (accessed 26 February 2021).

Geology of Britain. http://mapapps.bgs.ac.uk/geologyofbritain/home.html?&_ga=2.233755742.1734434781.1602596859-854914404.1598689897 (accessed 26 February 2021).

Indicative Flood Risk Areas for England. https://data.gov.uk/dataset/7792054a-068d-471b-8969-f53a22b0c9b2/indicative-flood-risk-areas-shapefiles (accessed 26 February 2021).

Map of runoff risk in Scotland (partial cover). https://www.spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/metadata/ab89fb3c-14fa-4ec1-8cd2-f9811a61713d (accessed 26 February 2021).

Map of soil erosion risk in Scotland (partial cover). https://www.spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/metadata/1a9302f1-b65e-40ae-8718-a4820cf70d1d (accessed 26 February 2021).

Maps of Soil compaction risk in Scotland (partial cover). https://www.spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/metadata/d16d695d-f7a0-4f1f-8ba2-335dda8428c8 (accessed 26 February 2021).

National Soil Inventory of Scotland (1978-88); National Soil Inventory of Scotland (2007-09); National Soil Map of Scotland 1:250000; Soil Map of Scotland (partial cover) 1:25000; Available Water Capacity Scotland: https://www.hutton.ac.uk/learning/natural-resource-datasets/soilshutton/soils-maps-scotland/download (accessed 26 February 2021).

National Soil Map of England and Wales. http://www.landis.org.uk/data/natmap.cfm (accessed 26 February 2021).

National Soils Inventory – Profile. http://www.landis.org.uk/data/nsiprofile.cfm (accessed 26 February 2021).

National Soils Inventory – Topsoil. http://www.landis.org.uk/data/nsitopsoil.cfm (accessed 26 February 2021).

Nitrate Vulnerable Zone designations 2017 to 2020 England. https://environment.data.gov.uk/farmers/ (accessed 26 February 2021).

Nitrate Vulnerable Zone designations 2017 to 2020 Scotland. https://map.environment.gov.scot/Soil_maps/?layer=9# (accessed 26 February 2021).

Nitrate Vulnerable Zone designations 2017 to 2020 Wales. http://lle.gov.wales/map#m=-3.70282,52.39942,8 (accessed 26 February 2021).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Northern Ireland Agricultural Census. https://www.daera-ni.gov.uk/publications/agricultural-census-northern-ireland-2019 (accessed 26 February 2021).

Scottish Pesticide Surveys Database. https://www.sasa.gov.uk/pesticides/pesticide-usage/scottish-pesticide-surveys-database-scopes(accessed 26 February 2021).

TELLUS Regional A Horizon Soils; TELLUS Regional Subsoil; TELLUS Regional A Horizon Soils-XRF. https://www.economy-ni.gov.uk/articles/introduction-tellus-project (accessed 26 February 2021).

The Crop Map of England. https://data.catchmentbasedapproach.org/datasets/crop-map-of-england-2018 (accessed 26 February 2021).

UK Climate Summary. https://www.metoffice.gov.uk/research/climate/maps-and-data/summaries/index (accessed 26 February 2021).

UK Soil and Herbage Pollutant Survey. https://www.gov.uk/government/publications/uk-soil-and-herbage-pollutant-survey (accessed 26 February 2021).

Water situation: national monthly reports for England. https://www.gov.uk/government/publications/water-situation-national-monthly-reports-for-england-2020 (accessed 26 February 2021).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

6.4 Chapter 3 – references

Albertoni, R., et al., 2014. 'EARTh: An Environmental Application Reference Thesaurus in the Linked Open Data Cloud. Semantic Web, vol. 5, no. 2, pp. 165-171. https://doi.org/10.3233/SW-130122

Keil, J.M., Schindler, S., 2019. Comparison and Evaluation of Ontologies for Units of Measurement. Semantic Web. 10 (1), 33-51. IOS Press. http://www.semantic-web-journal.net/system/files/swj1825.pdf

1.1 Turtle - Terse RDF Triple Language. World Wide Web Consortium (W3C).

Abdelbaki, A.M., 2016 Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils Ain Shams Eng. J. (2016), 10.1016/j.asej.2016.12.002

Al-Hashemi HMB, Al-Amoudi OSB, Yamani ZH, Mustafa YM, Ahmed H-u-R, 2021. The validity of laser diffraction system to reproduce hydrometer results for grain size analysis in geotechnical applications. PLoS ONE 16(1): e0245452. https://doi.org/10.1371/journal.pone.0245452

Al-Shammary, A.A.G., Kouzani, A.Z., Kaynak, A., Khoo, S.Y., Norton, M., Gates, W., 2018. Soil Bulk Density Estimation Methods: A Review, Pedosphere, 28 (4), 581-596, https://doi.org/10.1016/S1002-0160(18)60034-7.

Andrenelli, M.C., Fiori, V., Pellegrini, S., 2013. Soil particle-size analysis up to 250 μm by X-ray granulometer: device set-up and regressions for data conversion into pipette-equivalent values. Geoderma 192 (2013) 380–393. https://doi:10.1016/j.geoderma.2012.06.011

Arnaud, E., Laporte, M.A., Kim, S., Aubert, C., Leonelli, S., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P.L. and Mungall, C.J., 2020. The Ontologies Community of Practice: An Initiative by the CGIAR Platform for Big Data in Agriculture.

Atkinson, C. and K˙uhne, T. ,2003. Model-Driven Development: A MetamodelingFoundation.IEEE Software, 20(5):36–41.

Baize D., Girard M. (coords.), 2009. Référentiel pédologique 2008. Association française pour l'étude du sol (Afes). Éditions Quæ, Versailles, France. ISBN: 978-2-7592-0186-0. 435 pages.

Baker, T., Whitehead, B., Musker, R. et al., 2019. Global agricultural concept space: lightweight semantics for pragmatic interoperability. npj Sci Food 3, 16. https://doi.org/10.1038/s41538-019-0048-6

Baritz R, Erdogan H, Ahmadov H, Ghanma I, Lalljee VB, Wongmaneeroj A, Collins A, Monger C, Ribeiro JL, Bertsch F, Lalljee VB, with, Montanarella L, Comerma J, Khan A, VandenBygaart B, Gaistardo CC, Constantini E, Galbraith JM, Schad P, Lame F, Suvannang N, Hartmann C, Medyckyj-Scott D, Batjes NH, van Liedekerke M and Ziadat F 2017. Implementation Plan for Pillar Five of the Global Soil Partnership: Providing mechanisms for the collation, analysis and exchange of consistent and comparable global soil data and information, Rome, 48 p. http://www.fao.org/3/a-bs756e.pdf

Baritz R, Erdogan H, Fujii K, Takata Y, Nocita M, Bussian B, Batjes NH, Hempel J, Wilson P and Vargas R 2014. Harmonisation of methods, measurements and indicators for the sustainable management and protection of soil resources (Providing mechanisms for the collation, analysis and exchange of

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

consistent and comparable global soil data and information), Global Soil Partnership, FAO, 44 p. http://www.fao.org/3/a-az922e.pdf

Beckett, D., 2014. RDF 1.1 N-Triples - A line-based syntax for an RDF graph. World Wide Web Consortium (W3C).

Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G., 2014. RDF 1.1 Turtle - Terse RDF Triple Language. World Wide Web Consortium (W3C).

Bernard, S. A., 2012. An introduction to enterprise architecture. AuthorHouse.

Berners-Lee, T., 2006. Linked Data. Design Issues. World Wide Web Consortium (W3C).

Bittelli, M., Andrenelli, M.C., Simonetti, G., Pellegrini, S., Artioli, G., Piccoli, I., Morari, F., 2019. Shall we abandon sedimentation methods for particle size analysis in soils?, SOIL & TILLAGE RESEARCH, 2019, 185, pp. 36–46 https://doi.org/10.1016/j.still.2018.08.018

Bittelli, M., Flury, M., 2009. Errors in water retention curves determined with pressure plates. Soil Science Society of America Journal 73 (5), 1453-1460.

Booch, G,1993. Object-oriented Analysis and Design with Applications (2nd ed.). Redwood City: Benjamin Cummings. ISBN 0-8053-5340-2

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. Advances in Soil Science. 9. 177.

Bray, T., Paoli, J., Sperberg-McQueen, C. M., 1998. Extensible Markup Language (XML) 1.0. World Wide Web Consortium (W3C).

Brevik, E.C., Calzolari, C., Miller, B.A., Pereira P., Kabala C., Baumgarten, A., Jordán A., 2016. Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma, 264, pp. 256-274, 10.1016/j.geoderma.2015.05.017

Buchan, G. D., 1989. Applicability of the Simple Lognormal Model to Particle-size Distribution in Soils. Soil Science 147 (3): 155–161. doi:10.1097/00010694-198903000-00001

Bulens, J.D., M. Knotters, L.A.E. Vullings, D.J.J. Walvoort, H.R.M.J. Wehrens, 2021. Resolution and digital twins : Working with complex real-world phenomena in a digital environment. Wageningen University & Research. https://edepot.wur.nl/541412

Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S., Schaub, T., 2016. RFC 7946. IETF. doi:10.17487/RFC7946

Buttigieg, P.L., Pafilis, E., Lewis S.E., Schildhauer, M.P., Walls R.L., Mungall, C.J, 2016. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. J Biomed Semant, 7(1), 57. https://doi.org/10.1186/s13326-016-0097-6

Caracciolo, C., et al., 2013. The AGROVOC Linked Dataset. Semantic Web, vol. 4, no. 3, pp. 341-348. https://doi.org/10.3233/SW-130106

Chen, P., 1976. "The Entity-Relationship Model - Toward a Unified View of Data". ACM Transactions on Database Systems. 1 (1): 9–36. CiteSeerX 10.1.1.523.6679. doi:10.1145/320434.320440. S2CID 52801746.

Clastre, P., Pichot, C., Chanzy, A., 2018. Vocabulary for analysis and experimentation on terrestrial and aquatic ecosystems. Portail Data INRAE, V4. https://doi.org/10.15454/1.4894016754286177E12

Codd, E. F., 2002. A relational model of data for large shared data banks. In Software pioneers (pp. 263-294). Springer, Berlin, Heidelberg.

Costa, J.C., Borges, J.A.R., Pires, L.F., 2013. Soil bulk density evaluated by gamma-ray attenuation: Analysis of system geometry Soil and Tillage Research, 129, 23-31. https://doi.org/10.1016/j.still.2013.01.003

Cox, S.J.D., 2013. An explicit OWL representation of ISO/OGC Observations and Measurements. CEUR: 6th International Conference on Semantic Sensor Networks. URL: http://ceur-ws.org/Vol-1063/paper1.pdf

Cresswell, H. P., T. W. Green, and N. J. McKenzie, 2008. The adequacy of pressure plate apparatus for determining soil water retention. Soil Science Society of America Journal 72 (1), 41-49

Da Silva, A. R., 2015. Model-driven engineering: A survey supported by the unified conceptual model. Computer Languages, Systems & Structures, 43, 139-155.

Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., -, Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., Ashburner, M., 2008. ChEBI: a database and ontology for chemical entities of biological interest. D344–D350 Nucleic Acids Research, 2008, Vol. 36, Database issue Published online 11 October 2007. doi:10.1093/nar/gkm791

De Sousa, 2020. A prototype Ontology for GloSIS. Retrieved March 15, 2021. https://git.wur.nl/duque004/glosis-owl

Di Stefano, C., Ferro, V., Mirabile, S. , 2010. Comparison between grain-size analyses using laser diffraction and sedimentation methods. Biosystems Engineering 106(2):205-215. https://doi.org/10.1016/j.biosystemseng.2010.03.013EH&E, 2001. The precision and accuracy of environmental measurements for the building assessment survey and evaluation program. EH&E Report #11995, Environmental Health and Engineering Inc. (EH&E) for US Environmental Protection Agency, Washington DC, 25 p. http://www.epa.gov/iaq/base/pdfs/precisionandaccuracy.pdf

Du, H., Dimitrova, V., Magee, D., Stirling, R., Curioni, G., Reeves, H., Clarke, B.; Cohn, A., 2016 An ontology of soil properties and processes. In: Groth, Paul, (ed.) The semantic web – ISWC 2016. Springer, 30-37. (Lecture Notes in Computer Science, 9982). https://doi.org/10.1007/978-3-319-46547-0_4.

ECMA International, 2013. The JSON Data Interchange Format.

Ensmenger, N.L., 2009. The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise. MIT Press. p. 100. ISBN 978-0262050937. LCCN 2009052638

FAO – Unesco, 1974. Soil map of the world Volume 1, Legend, FAO Rome Italy, 59 pp.

FAO, 2006. Guidelines for soil description (4th Edn.), FAO, Rome, 97 pp.

FAO1988. Soil map of the world. Revised legend. World Soil Resources Report 60. FAO, Rome.

Feiden, K., Kruse F., Řezník T., Kubíček P., Schentz H., 2011. Best Practice Network GS SOIL Promoting Access to European, Interoperable and INSPIRE Compliant Soil Information. 9th International

Symposium on Environmental Software Systems (ISESS), Jun 2011, Brno,Czech Republic. pp.226-234, 10.1007/978-3-642-22285-6_25

Fisher P, Aumann C, Chia K, O'Halloran N, Chandra S (2017) Adequacy of laser diffraction for soil particle size analysis. PLoS ONE 12(5): e0176510. https://doi.org/10.1371/journal.pone.0176510.

Gilchrist, A. 2003. Thesauri, taxonomies and ontologies – an etymological note. Journal of documentation 59: 7–18

GlobalSoilMap, 2015. Specifications Tiered Global Soil Map products, Release 2.4 [07/12/2015)

Grigolon, G.B., Diotto, A.V., Gonçalves de Souza Lima, C.J., Francisco, J.P., Folegatti, M.V., 2020. Minimum number of points for a reliable soil water retention curve using Richard's pressure chamber. Revista Engenharia na Agricultura, 28, 477.487. doiI: 10.13083/reveng.v29i1.7320

Gruber, T., 1993. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". International Journal of Human-Computer Studies. 43 (5–6): 907–928. doi:10.1006/ijhc.1995.1081.

Haverkamp, R., Zammit, C., Bouraoui, F., Rajkai, K., Arrúe, J. L., and Heckmann, N., 1998. GRIZZLY, Grenoble Catalogue of Soils: Survey of soil field data and description of particle-size, soil water retention and hydraulic conductivity functions. Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE). France: Grenoble Cedex 9.

Hempel, J., Micheli, E., Owens, P.,McBratney, A., 2013. Universal Soil Classification System. Report from the International Union of Soil Sciences Working Group. Soil Horiz. 54, (2), 1–6.

Heon, M., 2020. G-OWL: Towards Graphical-Ontology-Web-Language-an OWL-2 Visual Notation for the Semantic Web Ontology Modeling.

Hodge, G. 2000. Systems of knowledge organization for digital libraries: beyond traditional authority files. Digital Library Federation. Washington, DC: The Digital Library Federation Council on Library and Information Resources

Hogg T.T.J., Henry J.L. 1984. Comparison of 1:1 and 1:2 suspensions and extracts with the saturation extracts in estimating salinity in Saskatchewan. Can. J. Soil Sci. 1984, 64, 699–704

Honderich, T., 2005a. "particulars and non-particulars". The Oxford Companion to Philosophy. Oxford University Press.Honderich, T. (2005b). "relations". The Oxford Companion to Philosophy. Oxford University Press.

ISO: International Standards Organisation,2011. ISO/IEC/IEEE 42010:2011 - Systems and software engineering - Architecture description.

ISO: International Standards Organisation. (2015). ISO 19150-2:2015 - Geographic information — Ontology — Part 2: Rules for developing ontologies in the Web Ontology Language (OWL).

IUSS Working Group WRB. 2015. World Reference Base for Soil Resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome. http://www.fao.org/3/i3794en/I3794en.pdf

Kargas G., Chatzigiakoumis I., Kollias A., Spiliotis D., Massas I., Kerkides P. 2018. Soil salinity assessment using saturated paste and mass soil:water 1:1 and 1:5 ratios extracts. Water, 10:1589, doi:10.3390/w10111589

Kendall, E. F., Bell, R., Burkhart, R., Dutra, M., & Wallace, E. K. (2009). Towards a Graphical Notation for OWL 2. In OWLED.

Klug H., Bretz B. (eds.), 2012. Discover INSPIRE compliant harmonised soil data and services. GS Soil Consortium, Paris-Lodron University Salzburg, Austria, 60 pages.

Kotsev A., Schleidt K., Liang S., van der Schaaf H., Khalafbeigi T., Grellet S., Lutz M., Jirka S., Beaufils M., 2018 - Extending INSPIRE to the Internet of Things through SensorThings API. Geosciences, MDPI, 2018,8 (6), pp.221. https://doi.org/10.3390/geosciences8060221, https://hal-brgm.archives-ouvertes.fr/hal-02003107

Krasilnikov, P., Ibáñez-Martí, J.J., Arnold, R., Shoba, S., 2009. A Handbook of Soil Terminology, Correlation, and Classification. Earthscan, London, UK.

L'Abate, G., Caracciolo, C., Pesce, V., Geser, G., Protonotarios, V., Costantini E.A.C. 2015. Exposing vocabularies for soil as Linked Open Data. Information Processing in Agriculture 2(3-4): 208-216. https://doi.org/10.1016/j.inpa.2015.10.002

L'Abate G., Caracciolo C., Villa F., Costantini E., 2016. Opening Soil Data through Semantics and Linked Vocabularies. Conferenza GARR 2016 - The Creactive Network. Uno spazio per condividere e creare nuova conoscenza – Abstracts. Firenze, 30 novembre-2 dicembre 2016. https://www.eventi.garr.it/it/conf16/home/materiali-conferenza-2016/paper/14-conf2016-paper-labate/file

L'Abate G., Caracciolo C., Villa F., Costantini E., 2017 Interoperabilità dei dati pedologici: dalla classificazione internazionale (WRB) ad una semantica formale. Conferenza GARR 2016 - The Creactive Network. Uno spazio per condividere e creare nuova conoscenza - Selected Papers. Firenze, 30 novembre-2 dicembre 2016 - ISBN: 978-88-905077-6-2 © 2017 Associazione Consortium GARR. DOI: 10.26314/GARR-Conf16-proceedings-07

Láng V., M. Fuchs, I. Waltner and E. Michéli, 2013. Soil taxonomic distance, a tool for correla-tion: As exemplified by the Hungarian Brown Forest Soils and related WRB Reference Soil Groups. Geoderma 192: 269-276.

Lankhorst, M.,2009. Enterprise architecture at work (Vol. 352). Berlin: Springer.

Lattelais C., Le Bas C., Schellenberger A., Yahiaoui R., 2021a. Publish data using SensorThings API. In progress.

Lattelais C., Le Bas C., Schellenberger A., Yahiaoui R., 2021b. Publish data through semantic web. Cookbook. EJP Soil. In progress.

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J., 2013. PROV-O: The PROV Ontology. (W3C Recommendation). World Wide Web Consortium. http://www.w3.org/TR/2013/REC-prov-o-20130430/

Leij, F. J., Alves, W. J., van Genuchten, M. Th., and Williams, J. R., 1996. Unsaturated Soil Hydraulic Database, UNSODA 1.0 User's Manual. Report EPA/600/R96/095. Ada, Oklahoma: US Environmental Agency, 103 pp.

Lohmann, S., Negru, S., Haag F., Ertl, T., 2016. Visualizing Ontologies with VOWL. Semantic Web 7(4): 399-419. WebVOWL - Web-based Visualization of Ontologies (visualdataweb.org)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Martini, D., Schmitz, M. and Mietzsch, E., 2013. agroRDF as a Semantic Overlay to agroXML: a General Model for Enhancing Interoperability in Agrifood Data Standards. In CIGR Conference on Sustainable Agriculture through ICT Innovation.

Mecum B., 2019. The Ecosystem Ontology. Retrieved March 15, 2021. https://bioportal.bioontology.org/ontologies/ECSO

Meersmans, J., Van Wesemael, B., and Van Molle, M. 2009. Determining soil organic carbon for agricultural soils. Soil Use Management, 25: 346– 353 https://doi.org/10.1111/j.1475-2743.2009.00242.x

Midford, P.E., Dececchi, T.A., Balhoff, J.P. et al., 2013. The vertebrate taxonomy ontology: a framework for reasoning across model organism and species phenotypes. J Biomed Semant 4, 34. https://doi.org/10.1186/2041-1480-4-34

Minasny, B., McBratney A.B., 2001. The Australian Soil Texture Boomerang: A Comparison of the Australian and USDA/FAO Soil Particle Classification Systems. Australian Journal of Soil Research 39 (6): 1443–1451. doi:10.1071/SR00065.

Nemes, A., Schaap, M. G., Leij, F. J., and Wösten, J. H. M., 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. Journal of Hydrology, 251(3–4), 151–162.

Nemes, A., Wosten J. H. M., Lilly A., Voshaar O. J. H., 1999. Evaluation of Different Procedures to Interpolate Particle-size Distributions to Achieve Compatibility within Soil Database. Geoderma 90 (3–4): 187–202.

OGC, 2011 - Observations and Measurements - XML Implementation. Version 2.0. http://www.opengis.net/doc/IS/OMXML/2.0. 77 pages

OGC: Open Geospatial Consortium, 2000. The OpenGIS Geography Markup Language Encoding Standard (GML).

OMG: Object Management Group, 2005. UML 2.0 Specification. http://www.omg.org/spec/UML/2.0/

Omuto C. T., Vargas R.R, El Mobarak, A. M., Mohamed N., Viatkin K., Yigini Y., 2020. Mapping of salt-affected soils: Technical manual. Rome, FAO https://doi.org/10.4060/ca9215en

Orilia, F.; Paolini Paoletti, M. (2020). "Properties". The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Ozcan H., Ekinci H., Yigini Y., Yuksel O., 2006. Comparison of four soil salinity extraction methods. Proceedings of 18th International Soil Meeting on “Soil Sustaining Life on Earth, Managing Soil and Technology”, May 22–26, 2006, Sanlıurfa, Turkey, p. 697–703

Palma, R., Janiak, B., GLOSIS Ontology: Layer-Horizon module. Revision: 2.0. Retrieved from: http://w3id.org/glosis/model/layerhorizon. Glosis Ontology (rapw3k.github.io)

Pieterse V., and Kourie, D. G. 2014. Lists, Taxonomies, Lattices, Thesauri and Ontologies. Knowl. Org. 2014 in: Org. 41-3, 217-229

Powers, D., 1991. Preface: Goals, Issues and Directions in Machine Learning of Natural Language and Ontology. AAAI Spring Symposium on Machine Learning of Natural Language and Ontology. DFKI.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

Rahman, S. A., H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi and M. Guizani, 2020. "A Survey on Federated Learning: The Journey from Centralized to Distributed On-Site Learning and Beyond," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2020.3030072.

Ravinder Kaur, Sanjeev Kumar and H. P. Gurung. , 2002. A pedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Australian Journal of Soil Research 40(5) 847 – 858

Rijgersberg, H., van Assem, M., Top, J. L., 2013. Ontology of Units of Measure and Related Concepts. Semantic Web, 4(1), 3-13. https://doi.org/10.3233/SW-2012-0069

Rijgersberg, H., Willems, D., Top J.L., 2021. Ontology of units of Measure (OM). Version: 2.0.29. Retrieved March 15, 2021, from https://www.ebi.ac.uk/ols/ontologies/om.

Ritchie, A.B.H., ed., 2016. OGC Soil Data Interoperability Experiment. Open Geospatial Consortium, Public Engineering Report, OGC 16-088r1. http://www.opengis.net/doc/PER/OGC-Soil-IE

Roussaki, I., Kosmides, P., Routis, G., Doolin, K., Pevtschin V., Marguglio, A., 2019. A Multi-Actor Approach to promote the employment of IoT in Agriculture. 2019 Global IoT Summit (GIoTS), Aarhus, Denmark, pp. 1-6, https://doi.org/10.1109/GIOTS.2019.8766416.

Rumbaugh, James, et al. Object-oriented modeling and design. Vol. 199. No. 1. Englewood Cliffs, NJ: Prentice-hall, 1991.

Schildhauer, M., . Jones, M. B., Bowers, S. Madin J.; Krivov S., Pennington, D., Villa, F., Leinfelder, B., Jones C., O'Brien, M., 2016. OBOE: the Extensible Observation Ontology, version 1.1. URL: http://dx.doi.org/10.5063/F11C1TTM

Selic, B., 2003. The Pragmatics of Model-Driven Development. IEEE Software, 20(5):19–25,

Semantic Sensor Network Incubator Group, 2011. Review of Sensor and Observations Ontologies. https://www.w3.org/2005/Incubator/ssn/wiki/Review_of_Sensor_and_Observations_Ontologies#O.26M-OWL_.28SemSOS.29

Shirazi, M. A., Boersma L., 1984. A Unifying Quantitative Analysis of Soil Texture. Soil Science Society of America Journal 48 (1): 142–147. doi:10.2136/sssaj1984.03615995004800010026x

Skaggs, T. H., Arya L. M., Shouse P. J., and Mohanty B. P., 2001. Estimating Particle-size Distribution from Limited Soil Texture Data. Soil Science Society of America Journal, 65 (4): 1038–1044. doi:10.2136/sssaj2001.6541038x.

Sleutel S., De Neve, S., Singier B., Hofman G., 2007. Quantification of Organic Carbon in Soils: A Comparison of Methodologies and Assessment of the Carbon Content of Organic Matter, Communications in Soil Science and Plant Analysis, 38:19-20, 2647-2657, DOI: 10.1080/00103620701662877

Soil Science Division Staff. 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Soil Survey Staff, 2012. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edition. USDA-NRCS, Washington, DC

Sonmez S., Buyuktas D., Asri F.O.. 2008.Assessment of different soil to water ratios (1:1, 1:2.5, 1:5) in soil salinity studies. Geoderma, 144: 361-369

Sporlein, P., Dilling, J., Joneck, M., 2004. Pilot study to test the equivalence or comparability of soil-particle-size analysis according to E DIN ISO 11277:06.94 (pipette method) and by the use of the SediGraph. Journal of Plant Nutrition and Soil Science 167, 649–656.

Sporny, M., Longley, D., Kellogg, G., Lanthaler, M. Lindström, N., 2019. JSON-LD Syntax 1.1. JSON for Linking Data W3C Community Group. .

Suvannang N. and Hartmann, C. 2019. First Inter-laboratory Comparison Report of the Regional Soil Laboratory Network for Asia (SEALNET). Rome, FAO.

Tahko, T. E.; Lowe, E. J., 2020. "Ontological Dependence" The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Takahashi T., Nakano K, Nira R, Kumagai E., Nishida M., Namikawa M., 2020. Conversion of soil particle size distribution and texture classification from ISSS system to FAO/USDA system in Japanese paddy soils, Soil Science and Plant Nutrition, 66:3, 407-414, DOI: 10.1080/00380768.2020.1763143

Tempel, P., Batjes, N. H., and van Engelen, V. W. P., 1996. IGBP-DIS soil data set for pedotransfer function development. Working paper and Preprint 96/05. Wageningen, The Netherlands: International Soil Reference and information Centre (ISRIC).

Thompson, H. S., Beech, D., Maloney, M., Mendelsohn N., 2001. XML Schema Part 1: Structures. World Wide Web Consortium (W3C).

Todd-Brown, K., Heran, L., Frederick, K., Younger M., Nave, L., van Gestel, N., Scheadel, C., Sierra, C., Lawrence, C., Hoyt, A., Stoner, S., Beem-Miller, J., Heckman, K., von Fromm, S., Kuhnen, Á., Holmquist, J., Wieder, W., Earl, S., Crowther, T., Batjes, N., Ribeiro, E.,2020. Soil Data Models: a review. 2020 November 23 Soil Ontology and Informatics Cluster. https://esipfed.github.io/soil_data_model_survey/index.html

van Genuchten, M Th, 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal, 44, 892–898.

Van Looy, K., Bouma, J., Herbst, M.,Koestel, J., Minasny, B., Mishra, U., … Vereecken, H. (2017). Pedotransfer functions in Earth system science: Challenges and perspectives. Reviews of Geophysics, 55, 1199–1256. https://doi.org/10.1002/2017RG0005

Vaz, C. M. P., and Hopmans, J. W., 2001. Simultaneous measurement of soil penetration resistance and water content with a combined penetrometer–TDR moisture probe. Soil Science Society of America Journal, 65, 4–12.

Wehrens, R., G. ten Broeke, H. Stigter, M. Gülpen, M. Knotters, M. Poortvliet, M. de Graaf, G. Hengeveld, 2021. Digital Twins for decision making, The uncertain half of the Twin, Wageningen University & Research report 3071, ISSN 1566-7197

Weynants, M. et al., 2013. European HYdropedological Data Inventory (EU-HYDI). JRC Technical Report, Report EUR 26053 EN. Luxembourg: Publications Office of the European Union, 169 pp. doi:10.2788/5936

Wilkinson, M., Dumontier, M., Aalbersberg, I.J. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wösten, J. H. M., Lilly, A., Nemes, A., and Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma, 90, 169–185

Zhang, Y., Schaap, M.G., 2019. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. Journal of Hydrology, 575, 1011-1030

Müller, H.W., Dohrmann, R., Klosa, D., Rehder, S., Eckelmann, W., 2009. Comparison of two procedures for particle-size analysis: Köhn pipette and X-ray granulometry. Journal of Plant Nutrition and Soil Science 172 (2), 172–179. http://dx.doi.org/10.1002/jpln.200800065

Nasta, P., Palladino, M., Sica, B., Pizzolante, A., Trifuoggi, M., Toscanesi, M., Giarra, A., D'Auria, J., Nicodemo, F., Mazzitelli, C., Lazzaro, Ul., Di Fiore, P., Romano, N., 2020. Evaluating pedotransfer functions for predicting soil bulk density using hierarchical mapping information in Campania, Italy, Geoderma Regional, 21., https://doi.org/10.1016/j.geodrs.2020.e00267https://doi.org/10.1016/j.geodrs.2020.e00267

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

6.5 Chapter 4 – references

Aktar, F., Bishop, T.F.A., Vervoort., R.W., 2021. Space-time modelling of groundwater level and salinity. Science of The Total Environment 776, 145865.

Box, G. E. P and Jenkins, G.M., 1976. Time series analysis: Forecasting and control, Holden-Day, San Francisco. Revised edition.

Arrouays, D., B.P. Marchant, N.P.A. Saby, J. Meersmans, T.G. Orton, M.P. Martin, P.H. Bellamy, R.M. Lark, M. Kibblewhite, 2012. Generic Issues on Broad-Scale Soil Monitoring Schemes: A Review. Pedosphere. 22(4): 456–469.

Brus, D.J., 1994. Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation. Geoderma 62, 233-246.

Brus, D.J., 2014, Statistical sampling approaches for soil monitoring, European Journal of Soil Science, 65, 779-791.

Brus, D.J., 2015. Balanced sampling: A versatile sampling approach for statistical soil surveys. Geoderma 253-254, 111-121.

Brus, D.J., 2019. Sampling for digital soil mapping: A tutorial supported by R scripts. Geoderma **338**: 464-480.

Brus, D.J. and de Gruijter, J. J., 1997, Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma, 80, 1-44.

Brus, D.J. and de Gruijter J.J., 2003. A method to combine non-probability sample data with probability sample data in estimating spatial means of environmental variables. Environmental Monitoring and Assessment 83: 303-317.

Brus, D.J., and de Gruijter, J.J., 2011, Design-based Generalized Least Squares estimation of status and trend of soil properties from monitoring data, Geoderma, 164, 172-180.

Brus, D.J. and de Gruijter, J.J., 2012, A hybrid design-based and model-based sampling approach to estimate the temporal trend of spatial means, Geoderma, 173-174, 241-248.

Brus, D.J., and de Gruijter, J.J., 2013, Effects of spatial pattern persistence on the performance of sampling designs for regional trend monitoring analyzed by simulation of space-time fields, Computers & Geosciences, 61, 175-183.

Brus, D.J., de Gruijter, J.J., van Groenigen, J.W., 2006a. Designing spatial coverage samples using k-means clustering algorithm. In: Lagacherie, P., McBratney, A.B., Voltz, M. (eds.) Digital soil mapping: An Introductory Perspective. Chapter 14, 183-192, Elsevier, Amsterdam.

Brus, D.J., de Gruijter, J.J. and de Vries, W., 2010, A sampling strategy for estimating plot average annual fluxes of chemical elements from forest soils, Geoderma, 159, 399-408.

Brus, D.J., B. Kempen, G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62, 394-407.

Brus, D.J. and Knotters M., 2008. Sampling design for compliance monitoring of surface water quality: A case in a Polder area, Water Resources Research, 44.

Brus, D.J., Knotters, M., Metzger, M.J. and Walvoort, D.J.J., 2011, Towards a European-wide sampling design for statistical monitoring of common habitats, Wageningen, Alterra, Alterra Report 2213

Brus, D.J., W. Nieuwenhuizen and A. Koomen, 2006b. Can we gain precision by sampling with probabilities proportional to size in surveying recent landscape changes in the Netherlands? *Environmental Monitoring and Assesment* **122**: 153-169.

Brus, D. J. and Saby, P. A., 2016. Approximating the variance of estimated means for systematic random sampling, illustrated with data of the French Soil Monitoring Network. Geoderma, 279, 77-86.

Brus, D.J., Spatjens, L.E.E.M., Gruijter de, J.J., 1999. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. Geoderma 89, 129-148.

Brus, D.J., W.J.M. Te Riele, 2001, Design-based regression estimators for spatial means of soil properties: the use of two-phase sampling when the means of the auxiliary variables are unknown. Geoderma 104. 257–279

Brus, D.J., 2021. Statistical approaches for spatial sample survey: Persistent misconceptions and new developments. European Journal of Soil Science 72, 686-703.

Brus, D.J., *in prep*. Spatial Sampling in R.

Cochran, W.G., 1977. Sampling techniques. Third edition. Wiley, New York.

Dalenius, T. & Hodges, J. L., Jr., 1959. Minimum variance stratification. Journal of the American Statistical Association, 54, 88–101

De Gruijter, J.J., D.J. Brus, M.F.P. Bierkens and M. Knotters, 2006. Sampling for natural resource monitoring. Springer, Berlin.

De Gruijter, J.J., Ter Braak, C.J.F., 1990. Model-Free Estimation from Spatial Samples: A Reappraisal of Classical Sampling Theory. Mathematical Geology 22(4), 407-415.

Gallego F.J., and Delincé, J. 2010. The European Land Use and Cover Area-frame statistical Survey (LUCAS). In Agricultural Survey Methods, R. Benedetti, M. Bee, G. Espa, F. Piersimoni. (Ed.), pp. 151-168 (New York: John Wiley & sons)

Grafström, A., Lundström, N., & Schelin, L., 2012. Spatially Balanced Sampling through the Pivotal Method. Biometrics, 68(2), 514-520. http://www.jstor.org/stable/23270453

Grafström, A., A. Hedström Ringvall, 2013. Improving forest field inventories by using remote sensing data in novel sampling designs. Canadian Journal of Forest Research 43 11. https://doi.org/10.1139/cjfr-2013-0123

Grafström, A., M. Ekström, B.G. Jonsson, P.-A. Esseen and G. Ståhl, 2019. On combining independent probability samples. Survey Methodology **45**(2): 349-364.

Van Groenigen, J.W. and Stein, A., 1998, Constrained Optimization of Spatial Sampling using Continuous Simulated Annealing, Journal of Environmental Quality, 27, 1078-1086.

Hartigan, J. A., 1975. Clustering Algorithms. Wiley Series in Probability and Mathematical Statistics.

Heuvelink, G.B.M., Griffith, D.A., Hengl, T. and Melles, S.J., 2012, Sampling Design Optimization for Space-Time Kriging, Spatio-Temporal Design: Advances in Efficient Data Acquisition, 9.

Heuvelink, G.B.M., Schoorl, J.M., Veldkamp, A. and Pennock, D.J., 2006, Space-time Kalman filtering of soil redistribution Geoderma, 133, 124-137.

Heuvelink, G.B.M., Griffith, D.A., Hengl., T., Melles, S.J., 2013. Sample design optimization for space-time kriging. In: Mateu, J., Muller, W.G., Spatio-temporal design: Advances in efficient data acquisition. John Wiley & Sons, Ltd.

Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-Hill, New York.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151, 331-326.

Lark, R.M., Ander, E.L. and Broadly, M.R., 2019, Combining two national-scale datasets to map soil properties , the case of available magnesium in England and Wales, European Journal of Soil Science 70, 361-377.

Lark, R.M. and Cullis, B.R., 2004, Model-based analysis using REML for inference from systematically sampled data on soil, European Journal of Soil Science,55, 799-813.

Laso Bayas, J.C., L. See, S. Fritz, T. Sturn, Chr. Perger, M. Dürauer, M. Karner, I. Moorthy, D. Schepaschenko, D. Domian and I. McCallum, 2016. Crowdsourcing in-situ data on land cover and land use using gamification and mobile technology. *Remote Sensing* **8**(905), doi:10.3390/rs8110905.

Lesch, S.M., Strauss, D.J. and Rhoades, J.D., 1995, Spatial Predicition of Soil Salinity Using Electromagnetic Induction Techniques: 2. An Efficient Spatial Sampling Algorithm Suitable for Multiple Linear Regession Model Identification and Estimation, Water Resources Research, 31. 378-398.

Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment* **132**: 202-211.

Maa, Tianwu, D.J. Brus, A-Xing Zhu, Lei Zhang, T. Scholten, 2020. Comparison of conditioned Latin hypercube and feature space coverage sampling for predicting soil classes using simulation from soil maps. Geoderma 370, 114-366, https://doi.org/10.1016/j.geoderma.2020.114366

Malone, B.P., Minansy, B. and Brungard, C., 2019, Some methods to improve the utility of conditioned Latin hypercube sampling. PeerJ 7:e6451 https://doi.org/10.7717/peerj.6451.

McBratney, A.B., Mendonça Santos, M.L,. Minasny, B., 2003. On digital soil mapping. Geoderma, 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Metzger, M.J. Bunce, R.G.H., Jongman, R.H.G, Mücher, C.A. and Watkins, J.W., 2005. A climatic stratification of the environment of Europe, Global Ecology and Biogeography, 14, 549-563

Minasny, B., McBrathey, A., 2006. A conditioned latin hypercube method for sampling in the presence of ancillary information. Computer and Geosciences 32(9), 1378-1388.

Morvan, X., N.P.A. Saby, D. Arrouays, C. Le Bas, R.J.A. Jones, F.G.A. Verheijen, P.H. Bellamy, M. Stephens, M.G. Kibblewhite. 2008. Soil monitoring in Europe: A review of existing systems and requirements for harmonisation. Science of the total environment 391, 1-12. https://doi.org:10.1016/j.scitotenv.2007.10.046

Overton, W.S. and Stehman, S.V., 1995, Design implications of anticipated data uses for comprehensive environmental monitoring programmes, Environmental and Ecological Statistics, 2 , 287-303.

Overton, W.S. and Stehman, S.V., 1996, Desirable design characteristics for long-term monitoring of ecological variables, Environmental and Ecological Statistics, 3, 349-361.

Roudier P, Beaudette D, Hewitt A. 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In: Minasny B, Malone B, McBratney AB, eds. Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. Leiden: CRC Press. 227-231

Stehman, S.V., C.C. Fonte, G.M. Foody and L. See, 2019. Using volunteered geographic information (VGI) in design-based statistical inference for area estimation and accuracy assessment of land cover. *Remote Sensing of Environment* **212**: 47-59.

Stoorvogel, J.J., Kempen, B., Heuvelink, G.B.M., de Bruin, S., 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. Geoderma 149, 161-170.

Thompson, S.K., 2002. *Sampling.* Second edition. Wiley, New York.

Tóth, G., Jones, A. and Montanarella, L, 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environmental Monitoring and Assessment 185:7409–7425

Truong, P.N., Heuvelink, G.B.M., Gosling, G.P., 2013. Web-based tool for expert elicitation of the variogram. Computers and Geosciences 51, 390-399.

van der Veeke, S., J. Limburg, R.L. Koomans, M. Soderstrom, S.N. de Waal, E.R. van der Graaf, 2021. Footprint and height corrections for UAV-borne gamma-ray spectrometry studies. Journal of Environmental Radioactivity 231, 106545, https://doi.org/10.1016/j.jenvrad.2021.106545

Van Groenigen, J.W., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. Journal of Environmental Quality 27, 1078-1086.

Vašát, R., Heuvelink, G.B.M., Borůvka, L., 2010. Sampling design optimization for multivariate soil mapping. Geoderma 155, 147-153.

Wadoux, A.M.J.C., Brus, D.J. and Heuvelink, G.B.M., 2019. Sampling design optimization for soil mapping with random forest, Geoderma, 355.

Walvoort, D.J.J., D.J. Brus and J.J. de Gruijter, 2010. An R-package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Computers & Geosciences 36(10): 1261-1267.

Webster, R., and Heuvelink, G.B.M., 2006. The Kalman filter for the pedologist's tool kit, European Journal of Soil Science, 57, 758-773.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Webster, R., M.A. Oliver, 2007. Geostatistics for Environmental Scientists, 2nd Edition. ISBN: 978-0-470-02858-2. 330 Pages

6.6 Chapter 5 – references

Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K., 2004. On-the-go soil sensors for precision agriculture. Comput. Electron. Agric. 44, 71–91.

Adamchuk, A., Ji, W., Viscarra Rossel, R.A., Gebbers, R., Tremblay, N., 2018. Chapter 9: Proximal Soil and Plant Sensing in Precision Agriculture Basics. D.K Shannon, D.E. Clay, and N. Kitchen (eds.) American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, 5585 Guilford Rd. Madison, WI 53711, USA.

Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. PLoS One, 9(8), e105519. DOI: 10.1371/journal.pone.0105519.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M.D., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P. A., Thompson, J. A., Zhang, G.-L., 2014. GlobalSoilMap: towards a fine-resolution global grid of soil properties. Advances in Agronomy, 125, 93-134.

Arrouays, D., Leenaars, J.G.B., Richer-de-Forges, A.C., Adhikari, K., Ballabio, C., Greve, M.H., Grundy, M., Guerrero, E., Hempel, J., Hengl, T., Heuvelink, G.B.M., Batjes, N.H., Carvalho, E., Hartemink, A.E., Hewitt, A., Hong, S.-Y., Krasilnikov, P., Lagacherie, P., Lelyk, G., Libohova, Z., Lilly, A., McBratney, A.B., McKenzie N.J., Vasques, G., Mulder, V.L., Minasny, B., Montanarella, L., Odeh, I.O.A,, Padarian, J., Poggio, L., Roudier, P., Saby, N.P.A,, Savin, I., Searle, R., Stolbovoy, V., Thompson, J.A., Smith, S., Sulaeman, Y., Vintila, R., Viscarra Rossel R.A., Wilson, P., Zhang, G.-L., ,Swerts M., van Oorts, K., Karklins, A., Feng, L., Ibelles Navarro, A.R., Levin, A., ,Laktionova T., Dell'Acqua, M., Suvannang, N., Ruam, W., Prasad, J., Patil, N., Husnjak, S., Pásztor, L., Okx, J., Hallet, S., Keay, C., Farewell, T., Lilja, H., Juilleret, J., Marx, S., Takata, Y., Kayusuki, Y., Mansuy, N., Panagos, P., van Liedekerke, M., Skalsky, R., Sobocka, J., Kobza, J., Eftekhari, K., Kazem Alavipanah, S., Moussadek,, R., Badraoui M., da Silva, M., Paterson, G., da Conceição Gonçalves, M., Theocharopoulos, S., Yemefack, M., Tedou, S., Vrscaj, B., Grob, U., Kozak, J., Boruvk,a L., Dobos, E., Taboada, M., Moretti, L., Rodriguez, D., 2017. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. GeoRes. J, 14, 1-19.

Arrouays, D., McBratney, A.B., Bouma, J., Libohova, Z., Richer-de-Forges, A.C., Morgan, C., Roudier, P., Poggio, L., Mulder V.L., 2020. Impressions of digital soil maps: the good, the not so good, and making them ever better. Geoderma Regional. 20, e00255. https://doi.org/10.1016/j.geodrs.2020.e00255

Bakacsi, Z., Laborczi, A., Szabó J., Takács, K., Pásztor, L., 2014. Proposed correlation between the legend of the 1:100.000 scale geological map and the FAO code system for soil parent material. Agrokémia és Talajtan, 63, 2,189–202. https://doi.org/10.1556/agrokem.63.2014.2.3.

Barber, J.J., Gelfand, A.E., Silander, J.A. Jr., 2006. Modelling map positional error to infer true feature location. Canadian Journal of Statistics, 34(4), 659-676.

Balkovič, J., Rampašeková, Z., Hutar, V., Sobocká, J., Skalský R. 2013. Digital Soil Mapping from Conventional Field Soil Observations. Soil and Water Research 8, 13-25. https://doi.org/10.17221/43/2012-SWR.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Batjes, N.H., Ribeiro, E., van Oostrum, A., 2020. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth Syst. Sci. Data, 12, 299–320. https://doi.org/10.5194/essd-12-299-2020.

Behrens, T., Schmidt, K., Macmillan, R.A., Viscarra Rossel, R.A.., 2018a. Multiscale contextual spatial modelling with the Gaussian scale space. Geoderma, 310, 128–137.Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan,, R.A. 2018b. Spatial modelling with Euclidean distance fields and machine learning. Eur. J. Soil Sci., 69, 757–770 doi: 10.1111/ejss.12687

Bethanney Janney, J., Emalda Roslin, S., Krishna Kumar, S., (2020) Analysis of skin lesions using machine learning techniques. Computational Intelligence and Its Applications in Healthcare, 73-90. https://doi.org/10.1016/B978-0-12-820604-1.00006-6

Bishop, T. F. A., McBratney, A. B., & Laslett, G. M. (1999). Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma, 91(1-2), 27-45. doi:10.1016/s0016-7061(99)00003-8

Bishop, T. F. A., Horta, A. & Karunaratne, S. B. (2015). Validation of digital soil maps at different spatial supports. Geoderma. 241, 238–249. https://doi.org/10.1016/j.geode rma.2014.11.026

Biswas, A. & Zhang, Y., 2018. Sampling designs for validating digital soil maps: A review. Pedosphere, 28(1), 1–15. https://doi. org/10.1016/S1002-0160(18)60001-3

Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M., Stum, A.K., 2008. Landsat Spectral Data for Digital Soil Mapping. In: Digital Soil Mapping With Limited Data. (Hartemink, A.E., McBratney, A.B., Mendonca Santos, M-d-L., Eds). 193-202. Springer, Dordrecht, The Netherlands.

Bonetto, R., Latzko, V., 2020. Chapter 8 - Machine learning. In Fitzek, F.H.P, Granelli, F., Seeling, P., (Eds.). Computing in Communication Networks. Academic Press, 2020-135-167. https://doi.org/10.1016/B978-0-12-820488-7.00021-9.

Borůvka, L., Penížek, V., Kozák, J., Marhoul, A.M., Vacek, O., Šrámek, V., Žížala, D., Poláková, Š., Smatanová, M., Fiala, P. (2018): Combining and harmonizing soil data from different sources: Problems and approaches. In: Arrouays, D., Savin, I., Leenaars, J., McBratney, A.B. (eds.): GlobalSoilMap. Digital Soil Mapping from Country to Globe. CRC Press/Balkema, Leiden, Netherlands, ISBN 978-0-8153-7548-7: pp. 29-34.

Bregt, A.K. 1992. Processing of soil survey data. Doctoral thesis. Wageningen Agricultural University, Wageningen.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. doi:Doi 10.1023/A:1010933404324

Brus, D. J., Kempen, B. & Heuvelink, G. B. M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science, 62(3), 394–407. https://doi. org/10.1111/j.1365-2389.2011.01364.x

Brus, D.J., Orton, T.G., Walvoort, D.J.J., Reijneveld, J.A., Oenema, O., 2014. Disaggregation of soil testing data on organic matter by the summary statistics approach to area-to-point kriging. Geoderma, 226-227, 151–159. https://doi.org/10.1016/j.geoderma.2014.02.011

Burrough P, McDonnell R (Eds). Principles of Geographical Information Systems. Oxford University Press, 1998.

Chen, S., Arrouays, D., Angers, D.A., Chenu, C., Barré, P., Martin, M.P., Saby, N.P.A., Walter, C., 2019a. National estimation of soil organic carbon storage potential for arable soils: a data-driven approach coupled with carbon-landscape zones. Sci. Tot. Env. 666, 355-367

Chen, S., Mulder, V.L., Martin, M.P., Walter, C., Lacoste, M., Richer-de-Forges, A.C., Saby, N.P.A., Loiseau, T., Hu, B., Arrouays, D., 2019b. Probability mapping of soil thickness by random survival forest at a national scale. Geoderma, 344, 184-194.

Chilés, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, Inc, New York.

Commission of the European Communities. 1985. Soil map of the European Communities 1:1000000. Directorate-General for Agriculture Coordination of Agricultural Research. EEC, Luxembourg.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8, 1991-2007, doi:10.5194/gmd-8-1991-2015

Cornell, R. and U. Schwertmann, 1996. The Iron Oxides. Structure, Properties, Reactions, Occurrence and Uses. VCH, Weinheim.

Cressie, N., Kornak, J., 2003. Spatial statistics in the presence of location error with an application to remote sensing of the environment. Statistical Science, 18(4), 436–456.

De Gruijter, J.J., D.J. Brus, M.F.P. Bierkens and M. Knotters, 2006. Sampling for natural resource monitoring. Springer, Berlin.

de Sousa, L. M., Poggio, L., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Riberio, E., and Rossiter, D. 2020: SoilGrids 2.0: producing quality-assessed soil information for the globe, SOIL Discuss. [preprint], https://doi.org/10.5194/soil-2020-65, in review

Dobos E, Carré F, Hengl T, Reuter HI, Tóth G. Digital Soil Mapping as a Support to Production of Functional Maps. 2006; EUR 22123 EN Office for Official Publications of the European Communities, Luxemburg.

Ellili Bargaoui, Y., , Walter, C., , Michot, D., Saby, N.P.A., Vincent, S., Lemercier, B., 2019. Validation of digital maps derived from spatial disaggregation of legacy soil maps. Geoderma 367, 113907. https://doi.org/10.1016/j.geoderma.2019.113907

Efron, B., 1992. Jackknife-after-bootstrap standard errors and influence functions. Journal of the Royal Statistical Society. Series B (Methodological), 83–127.

European Soil Bureau Network of the European Commission, 2005. Soil Atlas of Europe. A. Jones, L. Montanarella, and R. Jones (Eds) Office for Official Publications of the European Communities, Luxembourg. 2005. 128 p. ISBN 92 894 8120.

European Commission and the European Soil Bureau Network, 2004. The European Soil Database distribution version 2.0, , CD-ROM, EUR 19945 EN, 2004.

Evans, M. and F. Heller, 2003. Environmental Magnetism: Principles and Applications of Enviromagnetics. Academic Press.

Fantappiè, M., L'Abate, G., Costantini, E.A.C., 2010. Factors Influencing Soil Organic Carbon Stock Variations in Italy During the Last Three Decades. In Zdruli, P. et al. (eds.), Land Degradation and Desertification: Assessment, Mitigation and Remediation. Springer Science+Business Media. http://dx.doi.org/10.1007/978-90-481-8657-0_34

Fantappiè, M., L'Abate A., Costantini E.A.C., (2011). The influence of climate change on the soil organic carbon content in Italy from 1961 to 2008, Geomorphology (2011), http://dx.doi.org/10.1016/j.geomorph.2011.02.006FAO & ITPS, 2020. Global Soil Organic Carbon Map (v. 1.5): Technical Report. Rome, FAO.

FAO- Unesco (1971-1981). The FAO-Unesco soil map of the world. Legend and 9 vol. UNESCO, Paris

FAO-Unesco, 1981. Soil map of the world, volume V Europe, 1: 5 000 000, Rome.

Fernández-Ugalde O., Orgiazzi, A., Jones, A., Lugato, E., and Panagos, P., 2018. : LUCAS 2018 – SOIL COMPONENT: Sampling Instructions for Surveyors, ISBN 978-92-79-66255-3, ISSN 1831-9424, doi:10.2760/023673, Scientific and Technical Research Report, JRC105923, Publications Office of the European Union, 2017.

Fick, S. E. & Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas International Journal of Climatology, John Wiley & Sons, Ltd.

Finke, P., Hartwich, R., Dudal, R., Ibáñez, J.J., Jamagne, M., Kind, D., Montanarella, L., Yassoglou, N.J., 1998. Georeferenced Soil Database for Europe. Manual of Procedures. Version 1.0, 171.

Florinsky, I.V. Digital Terrain Analysis in Soil Science and Geology. Academic Press, Elsevier, 2012.

GlobalSoilMap, 2015. Specifications Tiered Global Soil Map products, Release 2.4 [07/12/2015]

Goovaerts P, 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Grandjean G., Cerdan, O., Richard, G., Cousin, I., Lagacherie, P., Tabbagh, A., Van Wesemael, B., Stevens, A., Lambot, S., Carré, F., Maftei, R., Hermann, T., Thörnelöf, M., Chiarantini, L., Moretti, S., McBratney, A.B., Ben Dor, E. (2010): DIGISOIL: An Integrated System of Data Collection Technologies for Mapping Soil Properties. In: Viscarra Rossel R., McBratney A., Minasny B. (eds.) Proximal Soil Sensing. Progress in Soil Science. Springer, Dordrecht, pp. 89-101. https://doi.org/10.1007/978-90-481-8859-8_7

Gray, J.M., Bishop, T.F.A., Wilford, J.R., 2016. Lithology and soil relationships for soil modelling and mapping. CATENA, 147, 429–440.

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. Geoderma, 143, 180-190.

Grunwald S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 2009: 152, 195-207. https://doi.org/10.1016/j.geoderma.2009.06.003

Hartemink, A.E., 2006. Soil maps of Europe. Book review of: Soil Atlas of Europe, by European Soil Bureau Network of the European Commission, 2005. Journal of Environmental Quality 35: 952–955.

Hartemink, A. E., 2008. Soil map density and nation's wealth and income. In: A.E. Hartemink, A.B. McBratney, M.L. Mendonca-Santos (Eds.), Digital Soil Mapping with Limited Data, Springer, Dordrecht (2008), pp. 53-66

Hartemink, A.E., Krasilnikov, P.V., Bockheim, J.G., 2013. Soil maps of the world. Geoderma, 207–208, 256–267. http://dx.doi.org/10.1016/j.geoderma.2013.05.003

Haykin, S., 2009. Neural Networks and Learning Machines. Pearson Education, Inc. ISBN-13: 978-0-13-147139-9

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present and future. Geoderma, 100, 269–301.

Heuvelink, G.B.M., Schoorl, J.M., Veldkamp, A. and Pennock, D.J., 2006, Space-time Kalman filtering of soil redistribution Geoderma, 133, 124-137.

Heuvelink, G.B.M., Griffith, D.A., Hengl., T., Melles, S.J., 2013. Sample design optimization for space-time kriging. In: Mateu, J., Muller, W.G., Spatio-temporal design: Advances in efficient data acquisition. John Wiley & Sons, Ltd.

Heuvelink,G., 2014. Uncertainty quantification of GlobalSoilMap products. In:GlobalSoilMap. CRC Press, pp. 335–340.https://doi.org/10.1201/b16500-62.

Helmick, J.L., Nauman, T.W., Thompson, J.A., 2014. Developing and assessing prediction intervals for soil property maps derived from legacy databases. In: GlobalSoilMap: basis of the global spatial soil information system (Arrouays, D., McKenzie, N.J., Hempel, J., Richer-de-Forges, A.C., McBratney, A.B., Eds).359-366. CRC press – Taylor & Francis, Boca Raton, FL, U.S.A.

Hengl T, Heuvelink GBM, Rossiter DG, 2007. About regression-kriging: from equations to case studies. Computers & Geosciences 33(10): 1301-1315.

Hengl T, Heuvelink GBM, Stein A, 2003. Comparison of kriging with external drift and regression-kriging. Technical note, ITC. https://www.researchgate.net/publication/228961429_Comparison_of_kriging_with_external_drift_and_regression-kriging

Hengl, T., Mendes de Jesus, M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., et al. (2014) SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS ONE 9(8): e105992. doi:10.1371/journal.pone.0105992

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE 12(2): e0169748. https://doi.org/10.1371/journal.pone.0169748

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. Peer J., 6: e5518; DOI 10.7717/peerj.5518

Hengl T, MacMillan RA, 2019. Predictive Soil Mapping With R. OpenGeoHub foundation, Wageningen, the Netherlands, 370 pages, www.soilmapper.org, ISBN: 978-0-359-30635-0.

Hengl T, Reuter, HI, 2008. Geomorphometry. Developments in Soil Science, Vol. 33, Elsevier, Amsterdam.

Hengl T, Rossiter DG, Stein A, 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Australian Journal of Soil Research 41: 1403-1422.

Hodgson, J.M., 1991. Soil survey - a basis for European soil protection: Proceedings of the meeting of European Heads of Soil Survey, 11 to 13 December 1989, Silsoe, UK

Hudson, B.D. 1992. The Soil Survey as paradigm-based Science. Soil Science Society of America Journal. https://10.2136/sssaj1992.03615995005600030027x.

Huggett, R.J., 1998. Soil chronosequences, soil development, and soil evolution: a critical review. Catena 32, 155–172. https://doi.org/10.1016/S0341-8162(98)00053-8

Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-Hill, New York.

Jones, R.J.A., Houšková, B., Bullock, P., Montanarella, L., 2005. Soil Resources of Europe, second edition.http://eusoils.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/esb_rr/n09_EUR20559.pdf (accessed 8 February 2021).

Kato, T., 2016. Chapter 4 - Prediction of photovoltaic power generation output and network operation. In Funabashi, T., Integration of Distributed Energy Resources in Power Systems. Implementation, Operation and Control, 77-108. https://doi.org/10.1016/B978-0-12-803212-1.00004-0

Knotters M, Brus DJ, Oude Voshaar JH, 1995. A comparison of kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma: 67(3-4), 227-246.

Knotters M. and D.J. Brus, 2013. Purposive versus random sampling for map validation: a cases study on ecotope maps of floodplains in the Netherlands. Ecohydrology 6: 425-434

Knotters, M., F.M. van Egmond, 2019, Structured Procedure for Selection of Suitable Soil Data Acquisition Techniques, presentation and short paper at 5th Global Workshop on Proximal Soil Sensing, Columbia, Missouri, USA

Knotters, M., G.B.M. Heuvelink, T. Hoogland & D.J.J. Walvoort, 2010. A disposition of interpolation techniques. Wageningen, Statutory Research Tasks Unit for Nature and the Environment, WOt-werkdocument 190

KristensenJ. A., Balstrom T., Jones, R.J.A., Jones A., Montanarella, L., Panagos, P., Breuning-Madsen, H., 2019. Development of a harmonised soil profile analytical database for Europe: a resource for supporting regional soil management. Soil, 5: 289–301

Kuhn M, Johnson K, 2013. Applied predictive modeling. Springer, New York, NY.INSPIRE Thematic Working Group Soil, 2013. D2.8.III.3 Data Specification on Soil – Technical Guidelines, D2.8.III.3_v3.0 (online) https://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SO_v3.0.pdf

Kyriakidis, P.C., 2004 A Geostatistical Framework for Area-to-Point Spatial Interpolation. Geographical Analysis 36: 259-289 doi:10.1111/j.1538-4632.2004.tb01135.x

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Ivushkin, K., Bartholomeus, H., Bregt, A.K., Pulatov, A., Kempen, B., de Sousa, L., 2019. Global mapping of soil salinity change. Remote Sensing of Environment, 231, 111260.

Lacoste, M., V.L. Mulder, A.C. Richer-de-Forges, M.P. Martin, D. Arrouays, 2016. Evaluating large-extent spatial modeling approaches: A case study for soil depth for France. Geoderma Regional 7-2, p 137-152 https://doi.org/10.1016/j.geodrs.2016.02.006

Lagacherie, P., Robbez-Masson, J.M., Nguyen-The, N., Barthès, J.P., 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma, 101, 105-118.

Lesaffre, E., Lawson, A.B., 2012. Bayesian Biostatistics. John Wiley & Sons, UK 978-0-470-01823-1, pp. 16–17.

Lagacherie, P., Gomez, C., 2018. Vis-NIR-SWIR Remote Sensing Products as New Soil Data for Digital Soil Mapping, In: McBratney, A.B., Minasny, B., Stockmann, U., (Eds.), Pedometrics. Progress in Soil Science, Springer, 415-438.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M.P., Saby, N.P.A., 2019. How far can the uncertainty on a Digital Soil Map be known?: a numerical experiment using pseudo values of clay content obtained from Vis-SWIR Hyperspectral imagery. Geoderma, 337, 1320-1328.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Nkuba-Kasanda, L., 2020. Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: a numerical experiment using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. Geoderma, 375, 114503. https://doi.org/10.1016/j.geoderma.2020.114503.

Legros, J.P. 2006. Mapping of the Soil. Science Publishers, Plymouth.

Loiseau, T., Chen, S., Mulder, V.L., Román Dobarco, M., Richer-de-Forges, A.C., Lehmann, S., Bourennane, H., Saby, N.P.A., Martin, M.P., Vaudour, E., Gomez, C., Lagacherie, P., Arrouays, D., 2019. Satellite data integration for soil clay content modelling at a national scale. Int. J .Appl. Earth Obs. Geoinformation, 82, 101905. https://doi.org/10.1016/j.jag.2019.101905.

Loiseau, T., Richer-de-Forges, A.C., Martelet, G., Bialkowki, A., Nehlig, P., Arrouays, D., 2020. Could airborne gamma spectrometric data replace lithological maps as co-variates for digital soil mapping of topsoil particle-size distribution? A case study in Western France. Geoderma Regional, 22, e00295.

Loiseau, T., Arrouays, D., Richer-de-Forges, A.C., Lagacherie, P., Ducommun, C., Minasny, B. 2021. Density of soil observations in digital soil mapping: A study in the Mayenne region, France. Geoderma Regional, 24, e00358.

Lorenzetti, R., Barbetti, R., Fantappiè, M., L'Abate, G., Costantini, E.A.C., 2015. Comparing data mining and deterministic pedology to assess the frequency of WRB reference soil groups in the legend of small scale maps. Geoderma 237–238, 237–245. http://dx.doi.org/10.1016/j.geoderma.2014.09.006

Ma, Y., Minasny, B., McBratney, A., Poggio, L., Fajardo, M., 2021. Predicting soil properties in 3D: Should depth be a covariate? Geoderma, 383. doi:10.1016/j.geoderma.2020.114794

Malone, B. P., McBratney, A. B., Minasny, B., & Laslett, G. M. (2009). Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma, 154(1-2), 138-152. doi:10.1016/j.geoderma.2009.10.007

Mahmoudabadi, E., Karimi, A., Haghnia, G.H., Sepehr, A., 2017. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. Environ. Monit. Assess., 189: 500. DOI 10.1007/s10661-017-6197-7

Martelet, G., Drufin, S., Tourlière, B., Saby, N.P.A., Perrin, J., DeParis, J., Prognon, J.F., Jolivet, C., Ratié, C., Arrouays, D., 2013. Regional regolith parameters prediction using the proxy of airborne gamma ray spectrometry. Vadose Zone Journal, 12(4). 1-14. doi:10.2136/vzj2013.01.0003

McBratney, A.B., Mendonça Santos, M.L,. Minasny, B., 2003. On digital soil mapping. Geoderma, 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecological Modelling, 411, 1, 108815.

Minasny, B., McBratney, A.B., 2010. Methodologies for global soil mapping. In: Boettinger, J.L., Howell, D., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping—Bridging Research, Environmental Application, and Operation. Progress in Soil Science, Springer, Dordrecht.

Minasny, B., McBratney, A.B. 2016: Digital soil mapping: A brief history and some lessons. Geoderma. 264, 301-311.

Montanarella, L., Jones, R.J.A., Dusart, J., 2005. The European Soil Bureau Network. In: Soil Resources of Europe, second edition. R.J.A. Jones, B. Houšková, P. Bullock and L. Montanarella (eds). European Soil Bureau Research Report No.9, EUR 20559 EN, (2005), 420pp. Office for Official Publications of the European Communities, Luxembourg, 3-14

Møller, A. B., Beucher, A. M., Pouladi, N., & Greve, M. H. (2020). Oblique geographic coordinates as covariates for digital soil mapping. SOIL, 6(2), 269-289. doi:10.5194/soil-6-269-2020

Moore ID, Grayson RB, Ladson AR, 1991. Digital terrain modelling: A review of hydrological, geomorphological, and bio-logical applications. Hydrological Processes 5(1): 3-30.

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I. & Shearer, S. A. (2004). Map quality for ordinary kriging and inverse distance weighted interpolation. Soil Science Society of America Journal, 68(6), 2042–2047. https://doi. org/10.2136/sssaj 2004.2042

Malone, B. P., McBratney, A. B. & Minasny, B., 2013. Spatial scaling for digital soil mapping. Soil Science Society of America Journal 77(3), 890–902. https://doi.org/10.2136/sssaj 2012.0419

Minasny, B. A.B. McBratney and A.E. Hartemink, 2010. Global pedodiversity, taxonomic distance, and the World Reference Base. Geoderma 155: 132-139 https://doi.org/10.1016/j.geoderma.2009.04.024

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma, 162, 1-19.

Mulder, V.L., Lacoste, M., Richer de Forges, A.C., Arrouays, D., 2016. GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. Sci. Tot. Env., 573, 1352-1369.

Myers DE, 1982. Matrix Formulation of Cokriging. Mathematical Geology 14: 249-257.

Němeček J., Damaška J., Hraško J., Bedrna Z., Zuska V., Tomášek M., KalendaM. 1967.

Survey of agricultural soils in Czechoslovakia, Vol. 1. MZVŽ, Praha (in Czech).

Novotný, I., Vopravil, J., Kohoutová, L., Poruba, M., Papaj, V., Khel, T., Žigmund, I., Vašků, Z., Tomiška, Z., Koutná, R., Pacola, M., Novotný, J., Havelková, L., Brouček, J., Žížala, D., 2013. Metodika mapování a aktualizace bonitovaných půdně ekologických jednotek. Výzkumný ústav meliorací a ochrany půdy, Praha (in Czech).

Odeh IOA, McBratney AB, Chittleborough DJ, 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma: 67(3–4): 215-226.

Nachtergaele, F.O.; Velthuizen, H. van; Wiberg, D.; Batjes, N.H.; Dijkshoorn, J.A.; Engelen, V.W.P. van; Fischer, G.; Jones, A.; Montanarela, L.; Petri, M.; Prieler, S.; Teixeira, E.; Shi, X. 2010. Harmonized World Soil Database (HWSD), 19th World Congress of Soil Science, Soil Solutions for a Changing World 1 – 6 August 2010, Brisbane, Australia. Published on DVD

Odgers, N.P., Libohova, Z. & Thompson J.A. 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189: 153-163.

Odgers, N.D., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. Geoderma, 214, 91-100.

Padarian, J., Minasny, B., and McBratney, A. B. 2019. Using deep learning for digital soil mapping, Soil, 5, 79–89

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: a review aided by machine learning tools. SOIL, 6(1), 35-52. doi:10.5194/soil-6-35-2020

Pal, R., 2017. Chapter 4 - Validation methodologies. In Pal (Ed.) Predictive Modeling of Drug Sensitivity, 83-107. https://doi.org/10.1016/B978-0-12-805274-7.00004-X

Panagos, P., Jones, A., Bosco, C., Kumar, P.S.S. 2011. European digital archive on soil maps (EuDASM): preserving important soil data for public free access. Int. J. Digit. Earth 4, 434–443. doi:10.1080/17538947.2011.596580

Panagos, P., Van Liedekerke, M. , Jones, A. , Montanarella, L., 2012. European Soil Data Centre: response to European policy support and public data requirements. Land Use Policy, 29 (2), pp. 329-338Petrovsky, E. and B. Elwood, 1998. Magnetic monitoring of air-, land- and water pollution. Quaternary Climates, Environments and Magnetism. Maher, B. Thompson, R., Cambridge.

Pasztor, L., Laborczi, A., Takacs, K., Illes, G., Szabo, J., Szatmari, G., 2020. Progress in the elaboration of GSM conform DSM products and their functional utilization in Hungary. Geoderma Regional, 21, e00269. DOI: 10.1016/j.geodrs.2020.e00269.

Piikki, K., & Söderström, M. 2019. Digital soil mapping of arable land in Sweden–Validation of performance at multiple scales. Geoderma. 352, 342–352. https://doi.org/10.1016/j.geode rma.2017.10.049

Piikki, K., Wetterlind, J., Söderström, M., Stenberg, B., 2021a. Perspectives on validation in digital soil mapping of continuous attributes—A review. Soil Use Manage. 37, 7-21 https://doi.org/10.1111/sum.12694

Piikki K., Wetterlind J., Söderstrom M. and Stenberg B. 2021b. Metrics and Plots for Model Evaluation. valmetrics R package version 1.0.0

Platou, S.W., Nørr, A.H. and Madsen, H. B., 1989. Digitisation of the EC Soil Map. In: Agriculture: computerization of land use data. R.J.A Jones and B. Biagi (eds.). EUR 11151 EN, 12-24. Office for Official Publications of the European Communities, Luxembourg.

Poggio, L., Gimona, A., 2017a. 3D mapping of soil texture in Scotland. Geoderma Regional, 9, 5-16. DOI: 10.1016/j.geodrs.2016.11.003

Poggio, L., Gimona, A., 2017b. A. Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas The Science of the Total Environment 579, p 1094-1110

Poggio, L.; Simonetti, E. & Gimona, A. 2018. Enhancing the WorldClim data set for national and regional applications Science of The Total Environment, 625, 1628 - 1643

Poggio, L.; Lassauce, A. & Gimona, A. 2019. Modelling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region. Geoderma, 346, 63-74

Poggio, L.; Gimona, A.; Spezia, L. & Brewer, M. 2016. Bayesian spatial modelling of soil properties and their uncertainty: the example of soil organic matter in Scotland using R-INLA. Geoderma, 277, 69-82

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2018). Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Science Society of America Journal, 0(0), 0. doi:10.2136/sssaj2017.04.0122

Rossiter, D.G., Rong Zeng and Gan-Lin Zhang, 2017. Accounting for taxonomic distance in accuracy assessment of soil class predictions. Geoderma 292: 118-127. https://doi.org/10.1016/j.geoderma.2017.01.012

Samuel-Rosa, A., G.B.M. Heuvelink, G.M. Vasques, L.H.C. Anjos, 2015. Do more detailed environmental covariates deliver more accurate soil maps? Geoderma 243–244, https://doi.org/10.1016/j.geoderma.2014.12.017

Schmidt, K., Behrens, T., Daumann, J., Ramirez-Lopez, L., Werban, U., Dietrich, P. & Scholten, T., 2014. A comparison of calibration sampling schemes at the field scale. Geoderma. 232–234, 243–256.

Science Committee. 2015. Specifications. Tiered GlobalSoil-Map products. GlobalSoilMap.net.

Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Toomanian, N., Scholten, T., 2020. Synthetic resampling strategies and machine learning for digital soil mapping in Iran. European Journal of Soil Science, 71, 352–368. https://doi.org/10.1111/ejss.12893

Thompson, R. and F. Oldfield, 1986. Environmental Magnetism. Allen and Unwin, Winchester.

de Sousa, L.M., Poggio, L., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing quality-assessed soil information for the globe. SOIL Discuss., https://doi.org/10.5194/soil-2020-65, 2020. Preprint under review for SOIL (Last access, 03-10-2021).

Soil Survey Division Staff, 1993. Soil Survey Manual. United States Department of Agriculture Handbook No. 18. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcseprd1335011.pdf

Soil Science Division Staff, 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054261

Szatmári, G., Pásztor, L., 2019. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. Geoderma 337, 1329-1340.

Taalab, K.; Corstanje, R.; Zawadzka, J.; Mayr, T.; Whelan, M.; Hannam, J. & Creamer, R. 2015. On the application of Bayesian Networks in Digital Soil Mapping Geoderma, 259, 134 - 148

Tavares-Wahren, F., Julich, S., Nunes, J.P., Gonzalez-Pelayo, O., Hawtree, D., Feger, K.-H., Keizer, J.J. 2015. Combining digital soil mapping and hydrological modeling in a data scarce watershed in north-central Portugal. Geoderma 264, 350–362. doi:10.1016/j.geoderma.2015.08.023.

Tavernier, R. Dudal, R. Osmond, D. Moormann, F., .1963. Soil map of Europe — Carte des sols de L'Europe — Mapa de suelos de Europe. Working Party on Soil Classification and Survey of the European Commission on Agriculture, FAO

Theodoridis, S., 2020. Chapter 18 - Neural Networks and Deep Learning. In Theodoridis, S., (Ed.). Machine Learning (Second Edition). A Bayesian and Optimization Perspective. Academic Press 2020-901-1038. https://doi.org/10.1016/B978-0-12-818803-3.00030-1.

Thompson, J.A., Nauman, T.W., Odgers, N.P., Libohova, Z., Hempel, J.W., 2012. Harmonization of legacy soil maps in North America: Status, trends, and implications for digital soil mapping efforts, in: Digital Soil Assessments and Beyond - Proceedings of the Fifth Global Workshop on Digital Soil Mapping. pp. 97–102.

Ting K.M. (2011) Confusion Matrix. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_157

Tobler W, 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46: 234–240.

Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environ. Monit. Assess. DOI 10.1007/s10661-013-3109-3.

Vaysse, K., Lagacherie, P., 2015. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma, Regional. 4, 20-30.

Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma, 291, 55-64. doi:10.1016/j.geoderma.2016.12.017

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik, J.O. Skjemstad, 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75, https://doi.org/10.1016/j.geoderma.2005.03.007.

Vossen, P. and Meyer-Roux, J. (1995). Crop monitoring and yield forecasting activities of the MARS Project. In: European Land Information Systems for Agro-environmental Monitoring. D. King, R.J.A. Jones and A.J. Thomasson, A.J. (eds). EUR 16232 EN, p.11-29. Office for Official Publications of the European Communities, Luxembourg, Wadoux, A. M.J-C., 2019. Using deep learning for multivariate mapping of soil with quantified uncertainty Geoderma, 351, 59-70

Wadoux AMJC, Minasny B, McBratney AB, 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Science Reviews 210: 103359.

Wadoux, A.M.J-C., Heuvelink, G.B.M., de Bruin, S. and Brus, D.J. (2021a). Confusion in statistical validation of large-scale ecological mapping models (under review)

Wadoux, A. M.J-C., D.J. Brus & D.J.J. Walvoort, 2021b (in prep). An integrated approach for the evaluation of quantitative soil maps through Taylor and target diagrams. To be submitted to Geoderma.

Waltari, E., R. Schroeder, K. McDonald, R.P. Anderson, A. Carnaval, 2014. Bioclimatic variables derived from remote sensing: assessment and application for species distribution modelling. Methods in Ecology and Evolution, 5, 1033-1042

Walvoort, D. J. J., M. Knotters, F.M. van Egmond, 2019. Interpolatie, aggregatie en desaggregatie van ruimtelijke bodemgegevens in de Basisregistratie Ondergrond (BRO). (WOt-technical report; No. 167). Wettelijke Onderzoekstaken Natuur & Milieu. https://doi.org/10.18174/514483

Webster, R., and Heuvelink, G.B.M., 2006, The Kalman filter for the pedologist's tool kit, European Journal of Soil Science, 57, 758-773.

Werban, U., Behrens, T., Cassiani, G., Dietrich, P. (2010): iSOIL: an EU project to integrate geophysics, digital soil mapping, and soil science In: Viscarra Rossel, R.A., McBratney, A., Minasny, B. (Eds.) Proximal Soil Sensing. Progress in Soil Science. Springer, Dordrecht. pp. 103-110, http://dx.doi.org/10.1007/978-90-481-8859-8_8

Wehrens, R. H. Putter & L. Buydens, 2000. The bootstrap: a tutorial.

Chemometrics and Intelligent Laboratory Systems 54:35-52 doi:10.1016/S0169-7439(00)00102-7.

Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma 183, 124–142.

Wilson, J. P., & Gallant, J. C. (Eds.), Terrain Analysis: Principles and Applications. New York: Wiley. 2000.

Wysocki, D.A., Schoeneberger, P.J., LaGarry, H.E., 2005. Soil surveys: a window to the subsurface. Geoderma. 126, 167–180. https://doi:10.1016/j.geoderma.2004.11.012.

Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K. and Vargas, R. (eds). 2018. Soil Organic Carbon Mapping Cookbook2nd edition. Rome, FAO. 220 pp

Zhang, Yakun, Wenjun Ji, D.D. Saurette, T.H, Easher, Hongyi Li, Zhou Shi, V.I. Adamchuk, A. Biswas, 2020. Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging. Geoderma 366, https://doi.org/10.1016/j.geoderma.2020.114253.

Zhu, X. X.; Tuia, D.; Mou, L.; Xia, G. S.; Zhang, L.; Xu, F. & Fraundorfer, F. 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine, 5, 8-36

Annex 1 Projects, programs or platforms having generated or used soil information at European level or beyond, and forthcoming needs

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
BonaRes	Soil as a sustainable resource for the bioeconomy https://www.bonares.de/	Germany	Data centre merging data from soil research, long-term field experiments and soil data from the public sector	BonaRes Repository: repository and web platform for soil and agricultural science data (http://doi.org/10.17616/R31NJMVY) Map of Long-term Field Experiments in Germany (and neighboring countries)	
CIRCASA	Coordination of International Research Cooperation on soil CArbon Sequestration in Agriculture https://www.circasa-project.eu/ (CIRCASA, 2020)	22 countries in several continents	Knowledge Information System (KIS) to share and visualize soil data and metadata from existing repositories	Data referencing to existing repositories for soil-related data (a.o. LUCAS 2019 Topsoil data, SoilGrids, GSOC); Data cartography for visualisation.	A monitoring, reporting and verification framework (MRV) system of SOC stocks in agricultural soils; The following elements pose requirements on soil databases: - Long-term field experiments at benchmark sites - Shorter-term field experiments - SOC/GHG models or SOC stock change factors - Spatial data to drive models (climate, land cover, soil properties including topsoil SOC content) (AgMIP, soil property maps from GSP/ FAO, LUCAS)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
					- Activity data (field and farm, management, self-reporting by farmers) - Remote sensing of land use, phenology, farmers' activities, addressing gaps in declarations to the CAP - Remote sensing to verify activity data, soils and vegetation inputs to run models - Spatial soil re-sampling survey grids (common grids for LUCAS and national monitoring systems - Harmonized data from EU Member States - Crowd-sourcing farm level data - Soil analysis by classical and proximal sensing methods (consistent methods in reference laboratory) Solutions capable of handling large volumes of data at high spatial resolution, using cloud technologies; An open-access database where geo-referenced, temporal soil C measurements can be uploaded,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
					standardized and shared (e.g. WoSIS, GloSIS); An online collaborative platform (e.g. as used in the CIRCASA project: https://www.ocp.circasa-project.eu/en/1/home.html)
Copernicus Land Monitoring Service	European system for monitoring the land surface using satellite imagery and in situ data https://land.copernicus.eu/	EEA39 countries	4 components: - Global Land Service - pan-European component - local component - imagery and reference data	Pan-European component: CORINE Land Cover datasets, High Resolution Layers and High Resolution Snow as Ice products; high-resolution phenology and productivity in 2021-2022 Local component: land cover and land use data in the Urban Atlas, Riparian Zones, Natura 2000 areas and Coastal Zones Imagery and reference data: EU-DEM and derived products, EU-Hydro (water bodies, drainage network and coastlines), LUCAS survey (land cover at observation points), European High Resolution Image mosaics	Issues with data access and restrictions on use of in-situ data managed at national level; Issues with data quality and availability across EEA39 countries.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
EBONE	European Biodiversity Observation Network https://cordis.europa.eu/project/id/212322/reporting https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Projects/EBONE.htm	Partners from 15 European countries, Israël and South Africa	System for collection of data on biodiversity at regional, national and European levels, with a focus on habitat information and linking field observations with Remote Sensing	European monitoring approach Selection of biodiversity indicators Habitat mapping and recording Habitat categories European Environmental Stratification, also applicable in agricultural research; A GIS database for European habitat and species data; Tested remote sensing approaches (LiDAR) for local habitat mapping and testing of phenology indicators	Recommendations on parameters on abiotic and biotic parameters of biodiversity to be collected at Long-term Ecosystem Research Sites (LTER) in Europe and remote sensing data, from both satellite and airborne data sources; Instructions for the mapping and recording of habitats; Soil data were not included in the Environmental Stratification of Europe, because soil data 'are not distinctive at the level of Europe, as the soil classification systems differ for each country.... and are not continuous but qualitative (class-based) data.'
ENVASSO	Environmental Assessment of Soil for Monitoring	25 EU Member States (in 2006)	Outline of a European Soil Monitoring System with a network of geo-referenced sites	60 indicators of soil threats selected based on thematic relevance, policy relevance and data availability;	A stepwise implementation is recommended, with the TOP3 indicators (which support minimum requirements) followed by a later extension using

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
	https://esdac.jrc.ec.europa.eu/projects/envasso			Three priority indicators for each threat are identified; Baseline and threshold values; Inventory of existing monitoring systems in the EU Member States (in 2006, red.) and of existing spatial assessments of soil threats; Listing of National Soil Monitoring Networks and gaps in coverage; Description of soil information systems in several Member States; Proposal for data model and soil database to provide an harmonised basis for capturing new soil data (including asking for standard measurement methods);	ENVASSO candidate indicators to achieve comprehensive soil monitoring; Several degradation types were not covered as diffuse contamination by organic contaminants, landslides; Considerable efforts were suggested to reach a common and acceptable standard of soil monitoring in Europe (number of monitoring sites to be installed, 1 site per 300 km²); A prototype database system has been constructed to provide web-based interoperability that is considered critical for the successful delivery of a comprehensive database holding up-to-date information and supporting access for the widest user community.

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
				22 procedures for indicators evaluated in 28 Pilot Areas	
ENVRI-FAIR	Environmental research infrastructures building FAIR services for research, innovation and society https://envri.eu/home-envri-fair/	37 partners from 13 EU countries	Open access platform for interdisciplinary environmental research data in the European Research Area utilising the European Open Science Cloud (EOSC)	Common data standards and policies for data life cycle, cataloguing, curation, provenance and service provision in Environmental Research Infrastructures; thematic data services and tools from the RI catalogues connected to the EOSC catalogue of services, COPERNICUS, GEO and other end-users.	Tools to manage, document, provide, find, access, and use EO data are still underdeveloped owing to a combination of data complexity and data volumes; Capitalize on progress made in the subdomains of atmosphere, marine, solid earth, and biodiversity/terrestrial ecosystems), and strengthen interoperability amongst Research Infrastructures and subdomains.
ESA WORLDSOILS (2020-ongoing)	part of the European Space Agency's (ESA) Earth Observation Strategy 2040 https://www.isric.org/projects/esa-worldsoils	7 partners (company and research) from EU countries and one external support contractor (Israël)	Pre-operational global Earth Observation-Soil Monitoring System for top soil organic carbon; Engaging end users to develop soil indices for monitoring topsoils;	Modular implementation to allow future extension to additional soil indices; Spatial resolution 100m x 100m globally and 50m x 50m over Europe; Appropriate confidence metrics;	Currently focuses on bare soils; Two approaches are followed, direct estimation of soil organic carbon from EO data and a digital soil mapping approach using EO data as covariates.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
				Large time series of >=3 years Validation over three European regions (Belgium, Czech Republic and Greece); Case studies with National Reference Centres of Soil.	
ESDAC	European Soil Data Centre https://esdac.jrc.ec.europa.eu/	principally EU and neighbouring countries, with additional data from other parts of the World	Data centre providing and merging data from soil surveys, monitoring, soil mapping and research projects etc. including data and metadata.	Data on soil properties, functions, threats and management information from regional to global scale.	
ESDB	European Soil Database v2 https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-	EU and neighbouring countries	Vector and attribute data and raster (grid) data files with cell sizes of both 1km x 1km and 10km x 10km for 73 soil related parameters, derived from the European Soil Database (ESDB) v2.0	4 discrete datasets: - the Soil Geographical Database of Eurasia at scale 1:1,000,000 (SGDBE) - the Pedotransfer Rules Database (PTRDB) - the Soil Profile Analytical Database of Europa (SPADBE)	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
	data#tabs-0-description=0			- the Database of Hydraulic Properties of European Soils (HYPRES)	
e-SOTER	Regional pilot platform as EU contribution to a Global Soil Observing System https://www.esoter.net/	14 partners from Europe, China, Morocco	A methodology to develop landform and parent material defined polygons at 1:1 M scale conform to the World Soil and Terrain database (SOTER) Methods to generate maps of soil and terrain attributes and digital data based on legacy soil survey data and remote sensing; A framework for cost-effective field survey and monitoring programs	e-SOTER made use of detailed digital elevation models (DEMs), remote sensing, and analytical tools for landform analysis, parent material detection and soil pattern recognition; Coverage of SOTER soil units for 4 windows at scale 1:1 M with harmonized soil classification and analytical soil data in a revised SOTER soil component data structure A 'cookbook' for installing the software tools necessary to develop and configure an e-SOTER web portal as a basis for project, national	Remote sensing cannot generate the soil patterns mapped in SOTER; Only a few remote-sensing techniques can penetrate below the ground surface; Incomplete coverage of e-SOTER units with input data (soil texture, parent material) hampered the use of models to assess soil threats; Limited access to national soil data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
				or regional datasets held in the SOTER format; Applications of the e-SOTER database for evaluating soil threats.	
EuDASM	European digital archive on soil maps https://esdac.jrc.ec.europa.eu/resource-type/national-soil-maps-eudasm https://www.isric.org/projects/eudasm-european-digital-archive-soil-maps		A collection of scanned soil maps (> 5400), largely derived from the historic ISRIC library collection, produced in the context of a project called European Digital Archive on Soil Maps of the World (EuDASM).		
GlobalSoilMap Working Group of IUSS (part of the GSP-P4 Working Group)	Arrouays et al., 2020	Countries participating in the GlobalSoilMap WG	Support to and specifications for the production of global soil property grids using approaches from countries to globe and vice versa	Maps and data on 12 soil properties at the support of pedon (point) and 100*100 m grid cells	Rescuing legacy data: methods for updating outdated data, including their uncertainty and for a better harmonization of laboratory measurements and soil classification;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
					To work on the uncertainty of indicators of prediction performances and to re-evaluate validation strategies; Explore and deploy the potential of high resolution remote sensing data for new covariates for mapping and monitoring soils; Improve the rational base behind the use of covariates (e.g. number, spatial resolution, pedological significance); Develop new covariates to capture historical changes in landscape and land use; Explore the use of other databases; Develop 4-D DSM by including temporal component

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
					Translate primary soil property products into information adjusted to end user needs through Digital Soil Assessment; Develop more intuitive metrics for uncertainty assessment for interpreting and evaluating soil maps; Incorporate pedological knowledge (e.g. on spatial patterns) into the map evaluation;
GS Soil	https://www.eurogeosurveys.org/projects/gssoil/	18 EU countries including 24 soil data providers	Spatial Data Infrastructure complementary to INSPIRE: GS Soil portal	Guidelines for harmonisation of soil information: reference terminology, commonly de-fined soil properties, common soil classification and harmonization of soil map geometries	GS Soil demonstrates the effort required to combine and harmonize soil maps from different sources throughout Europe under the INSPIRE architecture (Baritz et al., 2012). The GS Soil portal is not accessible.
GloSIS	Global Soil Information System http://www.fao.org/global-soil-partnership/areas-	Global Soil Partnership partners: national focal points of FAO member countries and	Spatial data infrastructure that links soil information collected by (national) institutions through a web-based platform	Four components: - Domain model - Data exchange standards - Node(s) connected to the internet and support node	Adoption of common standards for data exchange; The SDI for EJP SOIL should be connected to GLoSIS.

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
	of-work/soil-information-and-data/en/	international organisations. Supported by ISRIC – World Soil Information as GSP Soil Data Facility.		- Web-based gateway to GLOSIS nodes as access point for users: Discovery Hub Soil data products: soil profile and point data, global polygon coverage, global grids	
iSOIL	Interactions between soil related sciences - Linking geophysics, soil science and digital soil mapping https://esdac.jrc.ec.europa.eu/projects/isoil		Concepts and strategies for improved soil mapping through geophysical techniques, geophysical transfer functions and Digital Soil Mapping	Guidelines for collecting soil and auxiliary data for digital soil mapping and assessment, using geophysical, spectroscopic and monitoring techniques; Maps/datasets (geophysical, electromagnetic, spectroscopic) for 9 field sites in Germany, Czech Republic, Austria, Bulgaria.	
ISQAPER	Interactive Soil Quality Assessment https://www.isqaper-is.eu/	15 EU countries, Switzerland, China	Soil quality assessment app (SQAPP) for mobile devices, providing location-specific soil quality information and options for agricultural management practices (AMPs)	Spatial model of the potential for soil quality improvement and overall soil threat level, and ranking of AMPs; Pedoclimatic zonation for Europe, based on the SGDBE.	- link data from app to information on support for national greening measures; - enhance reporting functions to enable submission to ministries for compliance and monitoring; - a portal for viewing and downloading spatial information on potential for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
					improvement of soil quality, soil threat levels and rankings of AMPs; - enhanced collection and monitoring of data on soil quality, in particular on nutrients.
LANDMARK	LAND Management: Assessment, Research, Knowledge base http://landmark2020.eu/	13 EU countries	Soil monitoring scheme with harmonised indicators to enable assessment of five functions of agricultural soils in regions across Europe; Definition of required minimum set of data 3 for soil function modeling; Development of a decision support tool to assess the 5 soil functions (Soil navigator).	Regional indicators of soil functionality, tested at 100 sites across Europe, based on a matrix of soil type and land-use for the major climatic zones of Europe; Key indicators and minimum datasets for (i) primary production, (ii) water retention and regulation, (iii) habitat function and biodiversity, (iv) nutrient recycling and provision and (v) carbon cycling and sequestration); The Soil Navigator is able to calculate the 5 soil functions based on several measured or estimated parameters and provide threshold values to interpret the data.	Recommendation to collect information on land management in soil monitoring networks (e.g. irrigation, fertilization, pest control) and on environmental attributes (eg. topography, climatic properties) to assess the multifunctional capacity of a soil; Relevant information about horizons and biophysico-chemical analyses should be collected; A close collaboration with future LUCAS soil campaigns is encouraged.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
LANDSUPPORT	https://www.landsupport.eu/project/ (ongoing project started in 2018 for 4 years)	19 partners from 10 countries : Italy, Spain, Austria, Hungary, Germany; Belgium, France, Slovenia (includes Malaysia and Lebanon)	Aims at developing a free web-based, open-access GeoSpatial Decision Support System (S-DSS) devoted to: (i) support sustainable agriculture and forestry (ii) evaluate trade-off between land uses, and (iii) contribute to the development and implementation of land use policies in Europe.	This will require the integration of already existing databases (interoperability, including soil databases) at different coverage with the development of high performance modelling engines simulating agriculture & forestry (e.g. crop growth), land degradation and environmental issues (e.g. fate of pollutants, ecosystem services).	The project is ongoing and interactions with EJP SOIL-WP6 may be needed to agree on common formats for exchange of soil data and soil management.
LPIS	The Digital Land Parcel Information System of the Integrated Agricultural Control System (IACS) https://ec.europa.eu/jrc/en/research-topic/agricultural-monitoring	EU-27	LPIS are ICT and GIS instruments for EU Member State administrations to perform controls on payments for agricultural support payments allocated under the CAP.	Technical guidance and support to set up, refine and harmonise LPIS (https://marswiki.jrc.ec.europa.eu/wikicap/index.php/Main_Page); Image acquisition and storage for CAP Controls with Remote Sensing (CwRS) (satellite imagery and aerial photography); A LPIS Quality Assurance web portal is under development by JRC to reduce the need for on-the-spot controls;	Use of Remote Sensing and GIS in the implementation, management, and monitoring/control of the Good Agricultural and Environmental Conditions standard (GAEC)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
				DG AGRI uses the G4CAP web application to manage information flows for the CwRS and QA campaigns for LPIS between Member State Administrations, the JRC and DG AGRI and image providers, operators and contractors.	
LUCAS soil data	Land Use and Coverage Area frame Survey (LUCAS) https://esdac.jrc.ec.europa.eu/projects/lucas	EU countries and Albania, Bosnia-Herzegovina, Croatia, North Macedonia, Montenegro, Serbia and Switzerland	Information on the status of and changes in land use and land cover from >250.000 fixed sampling points every 3 years since 2006	LUCAS Soil: - harmonized and open-access dataset of soil properties, derived from samples collected at >23.000 fixed points and 0-20 cm depth, analysed in a central laboratory using standard methods on chemical, physical and biological parameters - derived maps - hosted by the European Soil Data Centre (ESDAC)	Issues found in the evaluation of LUCAS Soil Survey Laboratory Data (Hiederer, 2020) with data compliance to specifications, conformity to data structures and value ranges, consistency between parameters and surveys; Change in data delivery format would ease some of the format issues found; Including the analysis of texture parameters in revisited sample locations would help to identify problems in the stability of repeated samples.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
MAGIC	Marginal lands for Growing Industrial Crops https://magic-h2020.eu/	EU-28 and Ukraine	Mapping methodology and database for current and future marginal land in Europe[205]. Spatially-explicit classification (Marginal Agro-Ecological Zonation (M-AEZ)), to assess options to grow industrial crops in Europe. Accessible through ESRI data viewer	6 clusters of indicators of areas with natural constraints (18 in total) adapted from the approach of JRC[206]; Additional descriptive characteristics: - Classification of marginal lands according to rural area types based on FARO typology - overlap with HNV farmland and soil threats	
MARS	Monitoring Agricultural ResourceS https://ec.europa.eu/jrc/en/mars	Europe, sub-Saharan Africa and other areas of the world	Technical support services to DG Agriculture and Member State administrations for agricultural monitoring	Components: - Control with remote sensing (CwRS), the Digital Land Parcel Identification System (LPIS) and parcel area measurement using Global Navigation Satellite System (GNSS) devices. - MARS Crop Yield Forecasting System (MCYFS)	

[205] Elbersen et al. (2018)

[206] van Oorschoven et al., 2014, Terres, et al., 2014

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
				- Biophysical Models Applications (BioMA) framework	
RECARE	Preventing and Remediating degradation of soils in Europe through Land Care http://www.recare-hub.eu/recare-project	25 EU countries, Norway, Iceland and Switzerland; 17 study sites	Standardized and harmonized methods to monitor and assess soil degradation trends across Europe	List of key indicators for soil threats with temporal and spatial scales and methods, models and procedures for measurement or estimation; Repository for spatial data on soil threats, soil degradation and remediation measures, hosted in the European Soil Data Centre (ESDAC); WOCAT/LADA/DESIRE mapping method for land degradation and SLM, based on expert assessment https://www.wocat.net/en/global-slm-database/land-management-mapping-tools/	**List of key indicators:** - lack of harmonization on which methods/models to use over which spatial and - temporal scale - definition of peat soils - calculation of CO_2 emissions from soil - lack of accurate SOC estimations - lack of measured data on subsoil compaction and salinization - inconsistency in data on land take due to different methodologies applied by countries - deficiency in analytical and sampling techniques for soil contamination - lack of standardized procedures for assessment of desertification - lack of a method to provide an overall measure of soil biological health The **data repository** in ESDAC is no longer accessible due to 'copyright issues';

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	Title	Countries involved	Outcomes with regard to soil data structures	Detailed outcomes	Needs expressed/ difficulties
					The **mapping method for degradation and SLM** requires adjustments for application to European conditions: - separation of soil compaction and soil sealing - clearer distinction between loss of soil organic matter and other soil nutrients - map 'validation' is problematic, because existing map data are often already used by experts to compile the WOCAT maps - mapped variables of land degradation and SLM are qualitative
SoDaH	SOils DAta Harmonization database https://lter.github.io/som-website/	186 sites in various continents, mostly from the US; for Europe sites include LTERs https://www.lter-europe.net/	Open-source synthesis of data on soil organic matter and soil carbon from research networks ('experimental manipulations', gradient studies and time series)	Tools built for SoDAH aim to facilitate and automate harmonization and synthesis of soil carbon data; Individual locations can contribute results from (long-term) experiments; Tools (code) for data harmonization, aggregation, visualization and analysis.	How to group measurements taken from individual study locations that include diverse sampling protocols, unique experimental designs, and measurements from multiple soil depths; Locations may include manipulative experiments, gradient studies, and time series of repeated measurements; this requires users to be familiar with data structures to address scientific questions;

Acronym of project/ program or platform	**Title**	**Countries involved**	**Outcomes with regard to soil data structures**	**Detailed outcomes**	**Needs expressed/ difficulties**
					-> a need is to develop additional utilities to align and aggregate datasets from individual sites and locations.
SOILCARE	SoilCare For profitable and sustainable crop production in Europe https://www.soilcare-project.eu/	13 EU countries and Norway, Switzerland and UK	Database for monitoring performance of soil-improving cropping systems in experiments at field or farm level	Database architecture based on modern open-source software; Import and query system accessible via a web interface; Common database enabling comparison and combination of data across sites; R-scripts for unified statistical analysis.	Complex database structure requiring data entry in pre-designed spreadsheets with protected structure; Designed for small data volumes; Effort required from database managers in screening and cleaning delivered input data from sites.
SoilGrids	A system for predictive soil mapping based on compilation of global soil (profile) data and environmental layers using machine learning or AI approaches	Global coverage, compiled by ISRIC – World Soil Information	Portal and web-services providing access to SoilGrids maps and underpinning soil point data (WoSIS)	Global soil property maps at six standard depth intervals at a spatial resolution of 250 m, created by DSM, following the GlobalSoilMap specifications; quantified prediction uncertainty as 5th and 95th percentiles;	Accuracy of SoilGrids layers is still limited and the variation explained by the models is between 30% and 70%; reasons: - low density of soil profile observations in some regions (incl. data sharing issues); - missing reliable proxies at fine resolution for soil forming factors (e.g. parent material);

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Acronym of project/ program or platform	**Title**	**Countries involved**	**Outcomes with regard to soil data structures**	**Detailed outcomes**	**Needs expressed/ difficulties**
	https://soilgrids.org			Soil classification maps with Reference Soil Groups of the WRB	- other DSM methods may provide more accurate results, but would require excessive computation.
WoSIS	World Soil Information Service https://www.isric.org/explore/wosis/accessing-wosis-derived-datasets	global (https://www.isric.org/explore/wosis/wosis-contributing-institutions-and-experts)	Uniform data model and data ingestion/cleansing/standardization workflows	Standardised sets of soil properties at point locations (>190.000, CC-BY); Provides consistent point data for SoilGrids predictive mapping.	There are still many soil geographical and soil taxonomic gaps in WoSIS; Access constraints to numerous datasets.

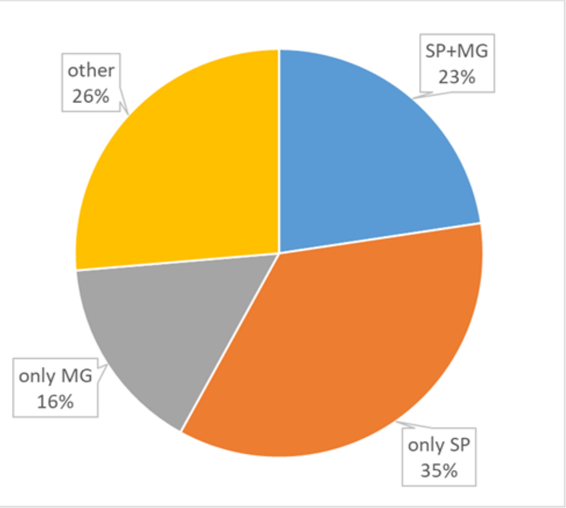
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Annex 2/A General description of data sources

Summary and distribution of data sources, reported by EJP SOIL partners, 2020

Where: “SP+MG”: data sources with soil property and management data; “only SP” is data source with only soil property data; “only MG” is data source with only soil management data, “other” means all the other reported data sources

	databases				
Country	**SP+MG**	**only SP**	**only MG**	**other**	**sum**
Austria	4	1	2	0	7
Belgium Flanders	0	4	4	8	16
Belgium Wallonia	3	1	0	0	4
Czech Republic	4	1	0	0	5
Denmark	2	4	1	0	7
Estonia	2	2	1	1	6
Finland	2	0	0	0	2
France	2	11	0	0	13
Germany	1	0	1	3	5
Hungary	1	2	2	0	5
Ireland	0	4	0	0	4
Italy	0	5	1	6	12
Latvia	9	6	0	0	15
Lithuania	3	0	0	0	3
Netherlands	4	8	0	0	12
Norway	0	5	2	8	15
Poland	1	5	0	0	6
Portugal	1	2	2	0	5
Slovakia	3	2	0	0	5
Slovenia	0	1	2	18	21
Spain	10	0	0	0	10
Sweden	3	6	1	5	15
Switzerland	0	1	0	0	1
Turkey	0	1	0	13	14
United Kingdom	0	14	19	2	35
sum	55	86	38	64	243
%	23	35	16	26	100



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

List of data sources (name, type, relevance), reported by EJP SOIL partners, 2020

Abbreviations and indications:

Data policy : F-freely ; R- freely for research purpose EJP ; P-permission; O- other
„as DS only”: the general database description is available without further details

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Abbreviation	Name	Data owner	SP	MG	as DS only	data policy	temporal relevance	type	spatial entity	geomety of mapping area
Austria	BORIS	Soil Information System	Umweltbundesamt GmbH/Environment Agency Austria	X	X	-	P	1980-2020	database	GRID	FULL
Austria	eBOD	Digital Soil Map of Austria	BFW - Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Lands	X	-	-	F	1950-2020	map	POINT	FULL
Austria	-	Österreichische Bodenschätzung (Austrian Soil Condition	Bundesminiterium für Finanzen/Ministry of Finance	X	X	-	P	1958-2020	database	POINT	FULL
Austria	-	Invekos - Agricultural Data	BMLRT - Federal Ministry of Agriculture, Regions and Tourism	-	X	-	P	-	data	-	FULL
Austria	AGES	AGES farm data and long term research sites	AGES-Austrian Agency for Health and FoodSafety	X	X	-	P	1988	database	POINT	OTHER
Austria	BAW	BAW specialized project data and long term research	BAW - Federal Agency for Water Management	X	X	-	P	1945-2020	database	POINT	OTHER
Austria	IfÖL	IfÖL - long term research site	IfÖL - Division of Organic Farming	-	X	-	O	2003-2020	network	GRID	NETWORK
Belgium F	DOV	DOV soil database for Flanders	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	X	-	-	F	1947-2020	database	POINT	FULL_REGIO
Belgium F	-	Soil map of Flanders (1:20.000)	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	X	-	-	F	1947-1978	map	NATURAL_POLYGONs	FULL_REGIO
Belgium F	-	Soil Organic Carbon Stock Maps for Belgium	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	X	-	-	F	-	map	GRID	FULL
Belgium F	-	Potential soil erosion map of Flanders	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	X	-	-	F	-	map	AGRI_PARCEL LPI	FULL_REGIO
Belgium F	-	Uitgevoerde gemeentelijke erosiebestrijdingswerken	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	-	X	-	F	-	map	ADMIN_POLYGONs	FULL_REGIO
Belgium F	-	WRB Soil units 40k	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	-	-	X	F	1947-1978	map	NATURAL_POLYGONs	FULL_REGIO
Belgium F	-	WRB Reference Soil Groups 250k	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	-	-	X	F	1947-1978	map	NATURAL_POLYGONs	FULL_REGIO
Belgium F	-	Bodemafdekkingskaart 2015 (BAK), 5 m resolutie (Soil sea	Vlaamse overheid, Agentschap Informatie Vlaanderen	-	-	X	F	2015	map	GRID	FULL_REGIO
Belgium F	-	Gevoeligheid voor grondverschuivingen (susceptilbity lar	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	-	-	X	F	2009	-	GRID	OTHER
Belgium F	-	Gekarteerde grondverschuivingen (mapped landslides)	Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving (VPO)	-	-	X	F	2009	-	NATURAL_POLYGONs	OTHER
Belgium F	-	Map with parcels under organic farming	-	-	-	X	P	-	map	-	FULL_REGIO
Belgium F	-	Map with the age of grassland (1, 2, 3, 4 and 5 years or old	-	-	-	X	P	-	map	-	FULL_REGIO
Belgium F	-	Bodemassociatiekaart (1:500.000)	Vlaamse overheid, Vlaamse Landmaatschappij	-	-	X	F	1972	map	NATURAL_POLYGONs	FULL
Belgium F	-	Bemestingsallocatie	Vlaamse overheid, Vlaamse Landmaatschappij	-	X	-	P	-	map	AGRI_PARCEL LPI	FULL
Belgium F	-	Landbouwgebruikspercelen	Vlaamse overheid, Departement Landbouw & Visserij	-	X	-	F	2008	map	AGRI_PARCEL LPI	FULL_REGIO
Belgium F	-	Groeicurve	Vlaamse overheid, Departement Landbouw & Visserij	-	X	-	F	2021	map	AGRI_PARCEL LPI	FULL_REGIO
Belgium W	CNSW	Digital Map of Walloon Soils	Public Service of Wallonia (SPW)	X	X	-	P	1947-1991	map	DERIVED	FULL_REGIO
Belgium W	AARDEWERK	Aardewerk database	Public Service of Wallonia (SPW)	X	-	-	P	1947-1991	database	POINT	FULL_REGIO
Belgium W	CARBIOSOL	Carbiosol map	Public Service of Wallonia (SPW)	X	X	-	P	1949-1972 2004-2014	map	GRID	FULL_REGIO
Belgium W	REQUASUD	REQUASUD database	REQUASUD asbl	X	X	-	P	1994-2020	database	POINT	FULL_REGIO
Czech Rep.	KPP	Komplexní průzkum půd (Systematic Soil Survey)	Ministry of Agriculture Czech Republic	X	X	-	O	1961-1971	database+map	POINT	TARGET AREAS
Czech Rep.	BPEJ	Bonitované půdně ekologické jednotky (Evaluated Soil Ec	State Land Office Czech Republic	X	-	-	F	1971-2020	map	NATURAL_POLYGONs	TARGET AREAS
Czech Rep.	LPIS	Land Registry LPIS (Land Parcel Identification System) - A	Ministry of Agriculture Czech Republic	X	X	-	O	-	database+map	FARMER_Block LPI	TARGET AREAS
Czech Rep.	-	Register of contaminated areas	Ministry of Agriculture Czech Republic	X	X	-	P	1990-2020	database	POINT	TARGET AREAS
Czech Rep.	-	Basal soil monitoring - basic	Ministry of Agriculture Czech Republic	X	X	-	P	1992-2020	database	POINT	TARGET AREAS
Denmark	DSPDB	Danish Soil Profile Database	Aarhus University	X	-	-	P	1980-2012	database	POINT	FULL
Denmark	DSCDB	Danish Soil Classification Data Base	Aarhus University	X	-	-	P	1975-1980	database	POINT	FULL
Denmark	DSMDB	Danish Soil Monitoring Data Base	Aarhus University	X	X	-	P	1986-2020 (2019)	database	POINT	FULL
Denmark	DASDB	Danish Acid Sulphate Data Base	Aarhus University	X	-	-	P	1981-1983	database	POINT	TARGET AREAS
Denmark	DSWDB	Danish Sinks Wetland Data Base	Aarhus University	X	-	-	P	2009-2010	database	POINT	TARGET AREAS
Denmark	DDSM	Danish Digital Soil Maps	Aarhus University	X	X	-	R	2014-2021	map	GRID	FULL
Denmark	DFR	Danish Farmers' Registrations	Aarhus University/Danish Agricultural Agency	-	X	-	P	2011-2020	map	FARMER_Block LPI	FULL
Estonia	PANDA	Eesti põllumuldade agrokeemiliste näitajate digitaalne ar	Agricultural Research Centre	X	-	-	O	2002-2020	database	POINT	TARGET AREAS
Estonia	Soil_map	Large-scale soil map of Estonia	Estonian Land Board	X	-	-	F	1954-1988 1997-2001	database+map	DERIVED	TARGET AREAS
Estonia	KESE	Riiklik mullaseire keskkonnaseire andmete kogumise ja a	Estonian Environment Agency	X	X	-	F	2002-2018	database	POINT	NETWORK
Estonia	-	Regular monitoring of arable soils (I period)	Estonian University of Life Sciences	X	X	-	R	1983-1994	database	POINT	NETWORK
Estonia	ARIB_register	The Agricultural Registers and Information Boards registe	The Agricultural Registers and Information Board (ARIB)	-	X	-	P	2004-2020	database	CATASDRAL_Parcel LPI	TARGET AREAS
Estonia	USLE_database	Soil erosion in Estonia modelled by use of USLE	Agricultural Research Centre (ARC)	-	-	X	P	2013, 2018	database	GRID	FULL

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Abbreviation	Name	Data owner	SP	MG	as DS only	data policy	temporal relevance	type	spatial entity	geometry of mapping area
Finland	VALSE	National monitoring of arable soil chemical quality 1974-2	-	X	X	-	P	1974-2018	database	-	-
Finland	VALSE subset	VALSE subset 2018	-	X	X	-	P	-	database	-	-
France	RMQS	French Soil Monitoring Network (Réseau de Mesures de l	Groupement d'Intérêt Scientifique SOL and INRAE	X	X	-	F	2001-2010	database	GRID	FULL
France	IGCS	Soil Management and Conservation Inventory	Groupement d'Intérêt Scientifique SOL and INRAE	X	-	-	P	1953-2020	database	NATURAL_POLYGONs	FULL
France	BDAT	Soil test database	Groupement d'Intérêt Scientifique SOL and INRAE	X	-	-	F	2013	database	NATURAL_POLYGONs	FULL_REGIO
France	BDGSF	Soil geographical database of France at 1:1,000,000	INRAE	X	X	-	F	1998	database	NATURAL_POLYGONs	FULL
France	-	*GlobalSoilmap France pH*	*GIS Sol and INRAE*	X	-	-	P	2014	database	GRID	FULL
France	-	*GlobalSoilmap France CEC*	*GIS Sol and INRAE*	X	-	-	P	2014	database	GRID	FULL
France	-	*GlobalSoilmap France depth*	*GIS Sol and INRAE*	X	-	-	P	2014	database	GRID	FULL
France	AWC BDGSF	Carte de la Réserve Utile en eau issue de la Base de Donn	INRAE	X	-	-	F	2018	database	NATURAL_POLYGONs	FULL
France	WRB BDGSF	Carte du nom de sol dominant en WRB issue de la Base de	INRAE	X	-	-	F	2017	database	NATURAL_POLYGONs	FULL
France	BDAT MAPS	Maps of the agronomic soil properties from BDAT	INRAE	X	-	-	P	2014	database	ADMIN_POLYGONs	FULL
France	RMQS MAPS ETM	Krigged maps of trace elements contents in French soil	INRAE	X	-	-	F	2014	database	GRID	FULL
France	-	*GlobalSoilmap France organic carbon*	*GIS Sol and INRAE*	X	-	-	P	2014	database	GRID	FULL
France	-	*GlobalSoilmap France available water content*	*GIS Sol and INRAE*	X	-	-	P	2014	database	GRID	FULL
Germany	BZE_LW	German Agricultural Soil Inventory	Thünen Institute	X	X	-	O	2012-2018	database	GRID	FULL
Germany	BUEK200	Soil map by the Federal Institute for Geosciences and Nat	Federal Institute for Geosciences and Natural Resources (BGR)	-	-	X	F	2020	database	NATURAL_POLYGONs	FULL
Germany	GUEK200	General Geological Map of the Federal Republic of Germa	Federal Institute for Geosciences and Natural Resources (BGR)	-	-	X	F	2020	database	NATURAL_POLYGONs	FULL_REGIO
Germany	DWD	Various interpolated grids characterizing weather and clir	Deutscher Wetterdienst	-	-	X	F	1881-2020	database	GRID	FULL_REGIO
Germany	DESTATIS	Landwirtschaftszählung/ Agrarstrukturerhebung	Statistisches Bundesamt (Destatis)	-	X		F	1930-2020	database	ADMIN_POLYGONs	FULL
Hungary	SIMS	Hungarian Soil Information and Monitoring System	Directorate of Plant Protection and Soil Conservation, National Food Chain Safety Office (NÉBIH)	X	-	-	P	1992-2010	database	POINT	OTHER
Hungary	DOSoReMi	Digital, Optimized, Soil Related Maps and Information in I	Institute for Soil Sciences, Centre for Agricultural Research (ATK TAKI)	X	-	-	P	-	database	GRID	FULL
Hungary	-	Nitrate Database	NEBIH (National Food Chain Safety Office)	-	X	-	P	2011-2020	database	AGRI_PARCEL LPI	TARGET AREAS
Hungary	MEPAR	Hungarian Identification System of Agricultural Parcels	National Center of Land Cases (Nemzeti Földügyi Központ)	X	X	-	P	2020	database	PHYSICAL_Block_LPI	FULL
Hungary	WEB_GN	Farmer Diary Program on the Web (Webes Gazdálkodási N	NEBIH (National Food Chain Safety Office)	-	X	-	P	2020	database	FARMER_Block LPI	NETWORK
Ireland	Irish SIS	Irish Soil Information System	Teagasc/EPA	X	-	-	O	2009-2014	database	DERIVED	FULL
Ireland	SQUARE	Soil Quality Assessment Research Project	Teagasc	X	-	-	P	2014-2015	database	OTHER	TARGET AREAS
Ireland	Tellus GSI	Tellus Geophyical and Geochemical Survey	GSI	X	-	-	F	2011-2023	database	GRID	FULL
Ireland	NSDB	National Soil Database of Ireland	Teagasc/EPA	X	-	-	P	1995 & 2003-2005	database	POINT	FULL
Italy	SISI-BADASUOLI	Italian Soil Information System	CREA	X	-	-	F	2013	database+map	DERIVED	FULL
Italy	GEO	Geological map of Italy by Compagnoni et al., 1983		-	-	X	F	1983	map	NATURAL_POLYGONs	FULL
Italy	SR	Italian Soil Regions	CREA	-	-	X	F	2013	-	NATURAL_POLYGONs	FULL_REGIO
Italy	Spectral_library	Spectral soil library (vis-NIR-MIR)	CREA	-	-	X	R	-	-	POINT	OTHER
Italy	TAIR	Long-term mean annual air temperature (1960-2008)	CREA	-	-	X	F	1960-2008	data	GRID	FULL
Italy	RICA	Rete di Informazione Contabile Agricola (Farms Agronomi	CREA	-	X	-	R	-	database	NATURAL_POLYGONs	TARGET AREAS
Italy	GSOCmap	*Global Soil Organic Carbon Map.*	*CREA-FAO-GSP*	-	-	X	F	1990-2013	-	GRID	FULL
Italy	-	Paper published data	Public	X	-	-	F	1953-2010	database	POINT	FULL
Italy	-	National surveys	CREA	X	-	-	F	1950-2020	database	POINT	FULL
Italy	-	Regional soil services	Regional Soil Services	X	-	-	P	1955-2018	database+map	POINT	FULL_REGIO
Italy	-	Private companies	Private companies	X	-	-	P	2008-2020	database+map	POINT	OTHER
Italy	-	CREA AA Italian historical wheater series	CREA	-	-	X	P	1961-2017	-	GRID	FULL_REGIO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Abbreviation	Name	Data owner	SP	MG	as DS only	data policy	temporal relevance	type	spatial entity	geomety of mapping area
Latvia	Digital profiles	Digital soil database (soil profiles)	Ministry of Agriculture of the Republic of Latvia	X	-	-	F	1960-1991	database	DERIVED	FULL
Latvia	Digital polygons	Digital soil database (soil polygons)	Ministry of Agriculture of the Republic of Latvia	X	-	-	F	1960-1991	database	POLYGON, without LPI	FULL
Latvia	LUCAS	*LV LUCAS Topsoil 2009 data*	*DG-Joint Research Centre, Euroean Union, EC*	X	-	-	F	2009	database	POINT	FULL
Latvia	GeoChem	Geochemical Atlas of Latvia	Ministry of Environmental Protection and Regional Development of the Republic Latvia	X	-	-	O	1997-2000	map	POINT	FULL
Latvia	SPPS	Soil agrochemical research database of the State Infor	State plant protection service	X	X	-	P	2004-2020	database	POINT	TARGET AREAS
Latvia	SPPS N	Database on the monitoring of mineral nitrogen in soi	State plant protection service	X	X	-	P	2005-2020	database	POINT	NETWORK
Latvia	SILAVA	National forest inventory	LSFRI Silava	X	-	-	P	2014-2018	-	POINT	FULL
Latvia	Pēterlauki	The long-term field trials "Pēterlauki"	Latvia University of Life Sciences and Technologies	X	X	-	O	1993-2020	paper	-	OTHER
Latvia	Vecauce	The long-term field trials "Vecauce"	Latvia University of Life Sciences and Technologies	X	X	-	O	1921-2020	paper	-	OTHER
Latvia	AREI Priekuļi	The long-term field trials in State Priekuļi Plant Breed	Institute of agricultural resources and economics	X	X	-	P	2001-2020	-	-	NETWORK
Latvia	AREI Stende	The long-term field trials in State Stende Cereals Bree	Institute of agricultural resources and economics	X	X	-	P	2001-2020	-	-	NETWORK
Latvia	ZMNĪ	Amelioration cadastre information system	State limited company "Ministry of Agriculture, Real Estate"	X	X	-	O	2018	map	GRID	FULL
Latvia	LLU Skrīveri	The long-term field trials "Sidrabiņi"	Latvia University of Life Sciences and Technologies	X	X	-	P	1981-2020	-	AGRI_PARCEL LPI	NETWORK
Latvia	LU ĢZZF	Universiy of Latvia fieldwork station	University of Latvia, Faculty of Geography and Earth Sciences	X	-	-	P	2019-2020	-	DERIVED	ADMIN_SMALL
Latvia	Dārzkopības institūts	The long-term field trials in the Institute of Horticultu	Latvia University of Life Sciences and Technologies	X	X	-	P	1945-2020	-	-	NETWORK
Lithuania	Dirv_DR10LT	Spatial Information Portal of Lithuania	National Land Service under the Ministry of Agriculture	X	X	-	F	1960-2020	database	POINT/POLYGON	TARGET AREAS
Lithuania	Dirv_vert	Spatial Information Portal of Lithuania	National Land Service under the Ministry of Agriculture	X	X	-	F	2019	database	POLYGON	TARGET AREAS
Lithuania	Mel_DR10LT	Spatial Information Portal of Lithuania	National Land Service under the Ministry of Agriculture	X	X	-	F	2020	database	POLYGON/LINE	TARGET AREAS
Netherlands	CC-NL	Carbon Change dataset of the Netherlands	WEnR	X	-	-	O	2018	database	POINT	FULL
Netherlands	SL_LTE_Networks	-	WEnR	X	X	-	O	-	database	-	NETWORK
Netherlands	BRO-BHR-p	Basisregistratie Ondergrond - Borehole research - prof	Ministry of the Interior and Kingdom Relations	X	-	-	F	1985-2020	description	POINT	FULL
Netherlands	BRO-SFR	Basisregistratie Ondergrond - Soil face research	Ministry of the Interior and Kingdom Relations	X	-	-	F	1985-2020	description	POINT	FULL
Netherlands	BRO-SGM	Basisregistratie Ondergrond - Soil map	Ministry of the Interior and Kingdom Relations	X	-	-	F	1970-2020	map	POLYGON, without LPI	FULL
Netherlands	PPS_BB - Bedrijvennet.	Publiek-Private Samenwerking Beter Bodembeheer -	Publiek-Private Samenwerking (PPS)	X	-	-	O	2019	-	POINT	NETWORK
Netherlands	PPS_BB - Bodemkw.	Publiek-Private Samenwerking Beter Bodembeheer -	Publiek-Private Samenwerking (PPS)	X	X	-	O	-	-	-	-
Netherlands	PPS_BB - iSQAPER	Publiek-Private Samenwerking Beter Bodembeheer -	Publiek-Private Samenwerking (PPS)	X	X	-	O	2016	-	-	-
Netherlands	PPS_BB - MAK	Publiek-Private Samenwerking Beter Bodembeheer -	Publiek-Private Samenwerking (PPS)	X	X	-	O	2016	-	-	-
Netherlands	BOFEK	Bodemfysische eenheden kaart	-	X	-	-	F	2012	map	-	-
Netherlands	BIS - Soilhydrophysics	Bodemfysische datasets BIS	Wageningen Environmental Research (WENR)	X	-	-	O	2012-2018	database	-	-
Netherlands	-	Geochemische atlas van Nederland (Geo-chemical alt	-	X	-	-	F	-	map	-	-
Norway	-	Norwegian soil survey sample grid 9 x9 km on agricultu	Norwegian Institute of Bioeconomy Research	X	-	-	F	1990-2020	database+map	NATURAL_POLYGONs	OTHER
Norway	-	Norwegian soil survey	Norwegian Institute of Bioeconomy Research	X	-	-	F	1990-2020	database+map	NATURAL_POLYGONs	TARGET AREAS
Norway	-	Norwegian soil property database	Norwegian Institute of Bioeconomy Research	X	-	-	P	1990-2020	database+map	POINT	TARGET AREAS
Norway	-	Observations and climate data as points	The Norwegian Water Resources and Energy Directorate; Norwegian Mapping Authority; The Norwegian Meteorological Institute	-	-	X	F	1957-2020	-	POINT	FULL
Norway	-	Observations and climate data as grid	The Norwegian Water Resources and Energy Directorate; Norwegian Mapping Authority; The Norwegian Meteorological Institute	-	-	X	F	1957-2020	-	GRID	FULL
Norway	AR5	Land resource map 1:5.000	Norwegian Institute of Bioeconomy Research	-	X	-	R	1960-2020	database+map	NATURAL_POLYGONs	TARGET AREAS
Norway	AR50	Land resource map 1:25.000	Norwegian Institute of Bioeconomy Research	-	-	X	F	1960-2020	map	NATURAL_POLYGONs	TARGET AREAS
Norway	DTM10	Digital terrain model 10x10 meters	Norwegian Mapping Authority;	-	-	X	F	2019	-	GRID	FULL
Norway	-	Bedrock, scale 1:50,000	Geological Survey of Norway	-	-	X	F	2004-2020	map	NATURAL_POLYGONs	FULL
Norway	Sediments	Superficial deposits 1:50 000 / 1:250.000	Geological Survey of Norway	-	-	X	F	2000-2020	map	NATURAL_POLYGONs	FULL
Norway	Marine Limit 1:50000	Marine limit (ML) indicates the highest level reached l	Geological Survey of Norway	-	-	X	F	2000-2020	map	NATURAL_POLYGONs	FULL
Norway	-	Avalanche and rockslide - rough susceptiblitity zones	The Norwegian Water Resources and Energy Directorate	X	-		F	2020	map	NATURAL_POLYGONs	FULL
Norway	-	Hydrology - mean annual runoff 1x1 km	The Norwegian Water Resources and Energy Directorate	-	-	X	F	1961-1990	-	GRID	FULL
Norway	-	Flood hazard maps are prepared for floods with 20-, 20	The Norwegian Water Resources and Energy Directorate	X	-	-	F	2020	map	NATURAL_POLYGONs	FULL
Norway	Direct Payments	Incentive Programs related to land use and animal hus	The Norwegian Agriculture Agency	-	X	-	F	1999-2020	database+map	POINT	FULL

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Abbreviation	Name	Data owner	SP	MG	as DS only	data policy	temporal relevance	type	spatial entity	geomety of mapping area
Poland	SAM25k	Soil Agricultural Map in scale 1:25000	IUNG	X	-	-	P	2013	map	POINT	FULL
Poland	ProfSAM25k	Referance soil profiles for SAM25	IUNG	X	-	-	P	1960-1980	-	POINT	FULL
Poland	Monit44k	Programme: "Assessment of ecological status of farmland i	IUNG	X	-	-	P	1992-1996	-	POINT	FULL
Poland	Monit41k		IUNG-MRiRW	X	-	-	P	2014-2015	-	POINT	FULL
Poland	MChGO	Monitoring of the Chemical Properties of Arable Soils in Po	IUNG-GIOŚ	X	-	-	F	1995-2015	-	POINT	TARGET AREAS
Poland	MonitFarm/Farm500		IUNG-MRiRW	X	-	-	R	2017-2018	-	POINT	NETWORK
Portugal	INFOSOLO	Database of Soil Profile Data for Portugal	INIAV	X	-	-	F	2014	database	POINT	OTHER
Portugal	PROPSOLO	Soil hydraulic properties of portuguese soils	INIAV	X	-	-	F	2011	database	POINT	OTHER
Portugal	-	*ICP Forests level I plots in Portugal sampled for the BioSoil s*	Instituto da Conservação da Natureza e das Florestas, IP (http://www.icnf.pt/) Contact person: Maria Barros (conceicao.barros@icnf.pt)	X	X	-	P	2008	-	GRID	OTHER
Portugal	-	COS 2018, 2018 Land Use and Soil Occupation Map for Portu	Direção-Geral do Território (DGT) (https://www.dgterritorio.gov.pt/Carta-de-Uso-e-Ocupacao-do-Solo-para-2018)	-	X	-	F	2018	map	OTHER	TARGET AREAS
Portugal	-	Public irrigated areas 2018 and SIR (Irrigation Information S	Direção-Geral de Agricultura e Desenvolvimento Rural (DGADR) as the National Irrigation Authority	-	X	-	F	2018	database+map	POINT	TARGET AREAS
Slovakia	BPEJ	Soil ecological units (land evaluation system)	NPPC	X	X	-	F	1971-2019	map	DERIVED	FULL
Slovakia	CMS - P	National soil monitoring system	NPPC	X	X	-	P	1993-2018	-	POINT	FULL
Slovakia	PM400	Soil Map of Slovakiain 1:400k scale	NPPC	X	-	-	F	1993	map	DERIVED	FULL
Slovakia	AISOP	Agricultural Soil Profiles Database	NPPC	X	-	-	P	1961-1970	database	POINT	FULL
Slovakia	LPIS	National Land Parcel Identification System	Ministry of Agriculture and Rural Development	X	X	-	P	2004-2020	-	PHYSICAL_Block_LPI	FULL
Slovenia	PEDOSEQUENCES	Pedosequences of Slovenia	Agricultural institute of Slovenia	-	-	X	O	1999	map	NATURAL_POLYGONs	FULL
Slovenia	SPM	Soil Pollution Monitoring	Ministry of Agriculture, Forestry and Food, Slovenia	-	-	X	O	1999 -	database	POINT	OTHER
Slovenia	SOIL MAP SLO	Soil Map of Slovenia	Ministry of Agriculture, Forestry and Food, Slovenia	-	-	X	F	1999	map	NATURAL_POLYGONs	FULL_REGIO
Slovenia	SOIL PROFILES SLO	Soil profiles of Slovenia	Ministry of Agriculture, Forestry and Food, Slovenia	X	-		F	2015	map	POINT	OTHER
Slovenia	SOIL QUAL INDEX	Soil quality index for Slovenia	Ministry of Agriculture, Forestry and Food, Slovenia	-	-	X	F	2006	map	NATURAL_POLYGONs	FULL
Slovenia	SOIL ORG MATTER	Soil organic matter in Slovenia	Agricultural institute of Slovenia	-	-	X	P	1999 - 2014	map	NATURAL_POLYGONs	TARGET AREAS
Slovenia	SOIL TEXTURE CLASS	Soil texture classes of Slovenia	Agricultural institute of Slovenia	-	-	X	P	1999 - 2014	map	NATURAL_POLYGONs	TARGET AREAS
Slovenia	SOIL PH	Soil pH for Slovenia	Agricultural institute of Slovenia	-	-	X	P	1999 - 2014	map	NATURAL_POLYGONs	TARGET AREAS
Slovenia	SOIL DEPTH	Soil depth for Slovenia	Agricultural institute of Slovenia	-	-	X	P	1999 - 2014	map	NATURAL_POLYGONs	TARGET AREAS
Slovenia	PLANT AV WATER	Plant available water in Slovenia	Agricultural institute of Slovenia	-	-	X	P	1999 - 2014	map	NATURAL_POLYGONs	FULL_REGIO
Slovenia	PROTECTED AREAS	Protected areas of Slovenia	Slovenian Environment Agency	-	-	X	F	2011 - 2011	map	NATURAL_POLYGONs	FULL
Slovenia	WATER PROT AREAS	Water protection areas of Slovenia	Slovenian Environment Agency	-	-	X	F	2011 - 2011	map	NATURAL_POLYGONs	FULL
Slovenia	FLOODING AREAS	Flooding areas of Slovenia	Ministry of the Environment and Spatial Planning - Slovenian Environment Agency	-	-	X	F	2013 - 2015	map	NATURAL_POLYGONs	FULL
Slovenia	LAND COVER	Soil cover of Slovenia	Ministry of the Environment and Spatial Planning - Slovenian Environment Agency	-	-	X	F	1996 - 2016	map	NATURAL_POLYGONs	FULL
Slovenia	LAND USE MAPS	Land use maps (Slovenia)	Ministry of Agriculture, Forestry and Food, Slovenia	-	-	X	F	2002 -	map	NATURAL_POLYGONs	FULL
Slovenia	LPIS	LPIS-parcels in Slovenia	Ministry of Agriculture, Forestry and Food, Slovenia	-	X	-	F	2005 -	map	NATURAL_POLYGONs	FULL
Slovenia	MELIORATION SYS	Melioration systems of Slovenia	Ministry of Agriculture, Forestry and Food, Slovenia	-	X	-	F	2013	map	AGRI_PARCEL LPI	FULL
Slovenia	DIGITAL ELEVATION	Digital elevation model of Slovenia	Surveying and Mapping Authority of the RePublic of Slovenia	-	-	X	F	2005 - 2005	map	GRID	FULL
Slovenia	GEOLOGICAL MAP	Basic geological map of Slovenia	Geological Survey of Slovenia	-	-	X	P	1969 - 1977	map	NATURAL_POLYGONs	FULL
Slovenia	LANDSLIDE PROB	Landslide probability map of Slovenia	Geological Survey of Slovenia	-	-	X	P	2005 - 2005	map	NATURAL_POLYGONs	FULL
Slovenia	FOREST COMP	Forest compartments in Slovenia	Slovenia Forest Service	-	-	X	F	1970-1990, 2017	map	NATURAL_POLYGONs	FULL

Country	Abbreviation	Name	Data owner	SP	MG	as DS only	data policy	temporal relevance	type	spatial entity	geomety of mapping area
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Soil	INIA	X	X	-	P	1995-2020	-	POINT	NETWORK
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Foo	INIA	X	X	-	P	1995-2020	-	POINT	NETWORK
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Eros	INIA	X	X	-	P	1995-2020	-	POINT	NETWORK
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Clim	INIA	X	X	-	P	1995-2020	-	POINT	NETWORK
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Indi	INIA	X	X	-	P	2005-2020	-	POINT	NETWORK
Spain	CEBAS Tres Caminos	CSIC-LTE CEBAS Tres Caminos	CSIC	X	X	-	P	2008-2020	-	POINT	NETWORK
Spain	CEBAS Cieza	CSIC-LTE CEBAS Cieza	CSIC	X	X	-	P	2006-2020	-	POINT	NETWORK
Spain	ICA La Poveda	CSIC-LTE ICA La Poveda	CSIC	X	X	-	P	2012-2020	-	POINT	NETWORK
Spain	Senés	CSIC-LTE EEAD Senes experimental farm	CSIC	X	X	-	P	2010	-	-	NETWORK
Spain	IRNASA	CSIC-LTE IRNASA Salamanca	-	X	X	-	P	-	-	-	NETWORK
Sweden	Soil & Crop Inventory	National Arable Soil & Crop Inventory	SEPA (Swedish Environmental Protection Agency)	X	-	-	P	1988-2017	database	POINT	FULL
Sweden	Arable Soils	National Inventory of Arable Soils	SBA (Swedish Board of Agriculture)	X	-	-	P	2011-2013	database	POINT	FULL
Sweden	Soil Compaction	National Soil Compaction Survey	SEPA (Swedish Environmental Protection Agency)	X	X	-	F	2003-2019	database	POINT	NETWORK
Sweden	Digital Soil Mapping	Digital Soil Mapping Sweden	SLU (Swedish University of Agricultural Sciences)	X	-	-	F	1965-2020	map	GRID	FULL
Sweden	Soil Types	Map of Soil Types	SGU (Swedish Geological Survey)	X	-	-	F	1960-2020	map	OTHER	FULL
Sweden	oring water from arable	Observationsfält på jordbruksmark (Monitoring water fro	SEPA (Swedish Environmental Protection Agency)	X	X	-	P	2005	database	POINT	NETWORK
Sweden	water from agricultural	Typområden på jordbruksmark (Monitoring water from ag	SEPA (Swedish Environmental Protection Agency)	X	X	-	P	1995-2011	database	POINT	NETWORK
Sweden	Weather data	Meteorologisk analysmodell MESAN (AROME) - API	SMHI	-	-	X	F	-	-	GRID	FULL
Sweden	NMD	National Land Cover Database (NMD)	SEPA (Swedish Environmental Protection Agency)	-	-	X	F	2017-2019	database	GRID	FULL
Sweden	gamma radiation measu	Geofysiska flygmätningar, gammastrålning (detaljerad) A	SGU (Swedish Geological Survey)	-	-	X	P	1968-	-	POINT	FULL
Sweden	Soil layer data	Jordlagerföljder (Soil layer data)	SGU (Swedish Geological Survey)	X	-	-	P	-	database	POINT	-
Sweden	Bedrock	GE.Berggrund 1:50 000 - 1:250 000	SGU (Swedish Geological Survey)	-	-	X	F	-	map	-	POINT
Sweden	Elevation	GSD-Elevation data, grid 2+	Lantmäteriet	-	-	X	P	-	-	GRID	FULL
Sweden	Arable filed data	Skiftesdata (Arable field data)	SBA (Swedish Board of Agriculture)	-	X	-	P	-	database+map	FARMER_Block LPI	FULL
Sweden	Erosion risk maps	Erosion risk maps	SBA (Swedish Board of Agriculture)	X	-	-	F	2018	map	GRID	OTHER
Switzerland	-	Swiss Soil Dataset Version 5 (2020)	NABODAT	X	-	-	F	1960-2016	database	POINT	OTHER
Turkey	GSOC Map	Global Soil Organic Carbon Map.	TAGEM-FAO-GSP	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	SOC Stock map	National Soil Carbon Stock Map	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	Boron Map	Boron Map of Turkey's Soils	TAGEM-National Boron Institute	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	Salinity Map	Salinity Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	pH Map	pH Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	K2O Map	K2O Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	P2O5 Map	P2O5 Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	CaCO3 Map	CacO3 Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	Sand Map	Sand Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	Silt Map	Silt Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	Clay Map	Clay Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	CEC Map	CEC Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	Bulk density Map	Bulk density Map of Turkey's Soils	TAGEM	-	-	X	P	2008-2009	map	POINT	FULL
Turkey	TAGEM_SFWRCRI_SIS	Turkish Ministry of Agriculture and Forestry Soil Fertilizer	TAGEM	X	-	-	P	-	database	POINT	FULL

Country	Abbreviation	Name	Data owner	SP	MG	as DS only	data policy	temporal relevance	type	spatial entity	geomety of mapping area
UK	NATMAP	National Soil Map of England and Wales	LandIS, University of Cranfield	X	-	-	P	1980-1995	map	NATURAL_POLYGONs	FULL
UK	NSI_Profile	National Soils Inventory - Profile	LandIS, University of Cranfield	X	-	-	P	1980-1995	database	POINT	FULL
UK	NSI_Topsoil	National Soils Inventory - Topsoil	LandIS, University of Cranfield	X	-	-	P	1983 & 1995	database	POINT	FULL
UK	CEH_Topsoil	Countryside survey of topsoil in Great Britain	UK Centre for Ecology & Hydrology	X	-	-	P	1978 & 1998 &2007	database	POINT	FULL
UK	CROME	The Crop Map of England	Catchment Based Approach (CaBA)	-	X	-	F	2018	map	GRID	FULL
UK	CEH_LC+Fert	CEH Land Cover plus Fertilisers	UK Centre for Ecology & Hydrology	-	X	-	P	2010-2015	map	GRID	FULL
UK	CEH_LC+Pest	CEH Land Cover plus Pesticides	UK Centre for Ecology & Hydrology	-	X	-	P	2012-2017	map	GRID	FULL
UK	CEH_LC+Crops	CEH Land Cover plus Crops	UK Centre for Ecology & Hydrology	-	X	-	P	2015-2019	map	GRID	FULL
UK	England_Water_Situation	Water situation: national monthly reports for England	Environment Agency - UK Government	-	X	-	F	2012-2020	map	-	FULL
UK	NVZ_England	Nitrate Vulnerable Zone designations 2017 to 2020	Environment Agency - UK Government	-	X	-	F	2017-2020	map	NATURAL_POLYGONs	FULL
UK	NVZ_Wales	Nitrate Vulnerable Zone designations 2017 to 2020	Natural Resources Wales	-	X	-	F	2020	map	NATURAL_POLYGONs	FULL
UK	GB_Bare_Soil_Erosion	Bare Soil Water Erosion Potential	British Geological Survey	-	X	-	F	1968-2013	map	GRID	FULL
UK	CEHLCM	CEH Land Cover Map 2017	UK Centre for Ecology & Hydrology	-	X	-	F	2017	map	GRID	FULL
UK	BGS_Geology	Geology of Britain	British Geological Survey	-	-	X	O	1968-2013	map	NATURAL_POLYGONs	FULL
UK	UK_Climate Summaries	UK Climate Summary	UK MetOffice	-	-	X	F	2001-2020	map	OTHER	FULL
UK	England_Flood_Risk	Indicative Flood Risk Areas for England	Environment Agency - UK Government	-	X	-	F	2020	map	GRID	FULL
UK	NSI_Scotland_1978-88	National Soil Inventory of Scotland (1978-88)	James Hutton Institute	X	-	-	F	1978-1988	database	POINT	FULL
UK	NSI_Scotland_2007-09	National Soil Inventory of Scotland (2007-09)	James Hutton Institute	X	-	-	F	2007-2009	database	POINT	FULL
UK	NSM_Scotland	National soil map of Scotland 1:250000	James Hutton Institute	X	-	-	F	1947-1981	database	NATURAL_POLYGONs	FULL
UK	NVZ_Scotland	Nitrate Vulnerable Zone designations 2017 to 2020	Scottish Government	-	X	-	F	2017-2020	map	NATURAL_POLYGONs	FULL
UK	NSM_Scotland_Partial	Soil map of Scotland (partial cover) 1:25000	James Hutton Institute	X	-	-	F	1947-1981	database	NATURAL_POLYGONs	OTHER
UK	AWC_Scotland	Available Water Capacity Scotland	James Hutton Institute	-	X	-	F	1947-1981	database	NATURAL_POLYGONs	FULL
UK	Erosion_Risk_Scotland	Map of soil erosion risk in Scotland (partial cover)	James Hutton Institute	-	X	-	F	1947-1981	map	NATURAL_POLYGONs	OTHER
UK	SCOPES	Scottish Pesticide Surveys Database	Scottish Government	-	X	-	F	2000-2020	database	-	FULL
UK	Runoff_Risk_Scotland	Map of runoff risk in Scotland (partial cover)	James Hutton Institute	-	X	-	F	1947-1981	map	NATURAL_POLYGONs	OTHER
UK	Compaction_Risk	Maps of Soil compaction risk in Scotland (partial cover)	James Hutton Institute	-	X	-	F	1947-1981	map	NATURAL_POLYGONs	OTHER
UK	NSM_Northern_Ireland	AFBI Soil Series Map of Northern Ireland	Agri-Food and Biosciences Institute	X	-	-	R	1988-1997	map	NATURAL_POLYGONs	FULL
UK	AFBI_5K_Attributes	AFBI 5k Soil Attribute Dataset	Agri-Food and Biosciences Institute	X	-	-	R	1988-1997	database	POINT	FULL
UK	UK_SHPS	UK Soil and Herbage Pollutant Survey	Environment Agency -UK Government	X	-	-	F	2007	database	POINT	FULL
UK	TELLUS_A	TELLUS Regional A Horizon Soils	Geological Survey of Northern Ireland	X	-	-	F	2004-2006	database	POINT	FULL
UK	TELLUS_B	TELLUS Regional Subsoil	Geological Survey of Northern Ireland	X	-	-	F	2004-2006	database	POINT	FULL
UK	TELLUS_XRF	TELLUS Regional A Horizon Soils-XRF	Geological Survey of Northern Ireland	X	-	-	F	2004-2006	database	POINT	FULL
UK	NI_Agri_Census	Northern Ireland Agricultural Census	Department of Agriculture Environment and Rural Affairs	-	X	-	F	2000-2019	database	-	-
UK	AFBI_Pesticide_Usage	AFBI Pesticide usage monitoring surveys	Agri-Food and Biosciences Institute	-	X	-	F	2000-2018	database	-	-
UK	Northern Ireland	The full area of Northern Ireland is identified as a NVZ	-	-	X	-	-	-	map	-	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Annex 2/B The thematic content of the databases for soil properties (SP), reported by the partners, supplemented with relevant ESDAC data

Abbreviations and indications:

SP1. Main soil properties (according to Global Soil Map specifications, 2015):

- SP1.1 Total profile depth
- SP1.2 Plant exploitable (effective) soil depth
- SP1.3 Organic carbon content
- SP1.4 pH in water
- SP1.5 Sand content
- SP1.6 Silt content
- SP1.7 Clay content
- SP1.8 Coarse fragments (gravel)
- SP1.9 ECEC
- SP1.10 Bulk density of the fine earth (< 2 mm) fraction *(excludes gravel)*
- SP1.11 Bulk density of the whole soil in situ *(includes gravel)*
- SP1.12 Available water capacity
- SP1.13 Electrical conductivity

SP2. Other soil properties:

- SP2.1 Calcium-carbonate content
- SP2.2 Field capacity (mm)
- SP2.3 pH KCl
- SP2.4 Saturated hydraulic conductivity *(Ksat)*
- SP2.5 Plant available amounts of macro and micro nutrients
- SP2.6 Total amounts of macro and micro nutrients/trace elements
- SP2.7 Quality of clay minerals
- SP2.8 Distribution of soil organisms

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

- SP2.9 Precompression stress
- SP2.10 Properties for NIR and MIR *(near and mid infrared)*

SP3. Threats (e.g. vulnerability)

- SP3.1 Soil erosion by water
- SP3.2 Soil erosion by wind
- SP3.3 Decline in SOM in peat soils
- SP3.4 Decline in SOM in mineral soils
- SP3.5 Compaction, structure degradation
- SP3.6 Soil sealing
- SP3.7 Pollution with potentially toxic elements
- SP3.8 Pollution with organic substances
- SP3.9 Acidification
- SP3.10 Salinization and alkalization
- SP3.11 Desertification
- SP3.12 Flooding
- SP3.13 Landslide
- SP3.14 Decline in soil biodiversity

SP4. Soil classification and fertility:

- SP4.1 Soil type, national classification
- SP4.2 Soil type, international classification
- SP4.3 Soil fertility
- SP4.4 Data for status of soil biodiversity/health

SP5. Soil functions/services:

- SP5.1 Water storage capacity for e.g. topsoil or until a given soil depth *(different from SP1.12 or SP2.2)*
- SP5.2 Plant productivity potential
- SP5.3 Carbon sequestration potential
- SP5.4 Biodiversity potential

* indicates reported SP dataset without further details

CountrY	Database		SP1													SP2										SP3														SP4				SP5			
	Abbreviation	Name	SP1.1	SP1.2	SP1.3	SP1.4	SP1.5	SP1.6	SP1.7	SP1.8	SP1.9	SP1.10	SP1.11	SP1.12	SP1.13	SP2.1	SP2.2	SP2.3	SP2.4	SP2.5	SP2.6	SP2.7	SP2.8	SP2.9	SP2.10	SP3.1	SP3.2	SP3.3	SP3.4	SP3.5	SP3.6	SP3.7	SP3.8	SP3.9	SP3.10	SP3.11	SP3.12	SP3.13	SP3.14	SP4.1	SP4.2	SP4.3	SP4.4	SP5.1	SP5.2	SP5.3	SP5.4
Austria	BORIS	Soil Information System	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	eBOD	Digital Soil Map of Austria	Y	-	Y	Y	Y	Y	Y	-	-	-	-	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Österreichische Bodenschätzung (Austrian Soil Condition Survey)	Y	-	-	-	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	AGES	AGES farm data and long term research sites	Y	-	Y	Y	Y	Y	Y	-	Y	-	-	-	Y	Y	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	BAW	BAW specialized project data and long term research	Y	-	Y	Y	Y	Y	Y	-	-	Y	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	Y	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Belgium Flanders	DOV	DOV soil database for Flanders	Y	-	Y	Y	Y	Y	Y	Y	-	-	Y	-	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	Soil map of Flanders (1:20.000)	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	Soil Organic Carbon Stock Maps for Belgium	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Potential soil erosion map of Flanders	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Belgium Wallonia	CNSW	Digital Map of Walloon Soils	Y	-	-	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	AARDEWERK	Aardewerk database	Y	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	CARBIOSOL	Carbiosol map	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	REQUASUD	REQUASUD database	-	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Czech Republic	KPP	Komplexní průzkum půd (SYstematic Soil Survey)	Y	-	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	BPEJ	Bonitované půdně ekologické jedNtkY (Evaluated Soil Ecological units)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-
	LPIS	Land RegistrY LPIS (Land Parcel Identification SYstem) - AZZP	-	-	Y	-	Y	Y	Y	-	Y	-	-	-	-	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Register of contaminated areas	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Basal soil monitoring - basic	-	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
Denmark	DSPDB	Danish Soil Profile Database	Y	-	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	-	Y	Y	Y	-	Y	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	DSCDB	Danish Soil Classification Data Base	-	-	Y	-	Y	Y	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DSMDB	Danish Soil Monitoring Data Base	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DASDB	Danish Acid Sulphate Data Base	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	DSWDB	Danish Sinks Wetland Data Base	-	-	Y	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DDSM	Danish Digital Soil Maps	-	-	Y	Y	Y	Y	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
Estonia	PANDA	Eesti põllumuldade agrokeemiliste näitajate digitaalne andmekogu (Digital database of agrochemical parameters of Estoanian agricultural soils)	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	Y	-
	Soil_map	Large-scale soil map of Estonia	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	KESE	Riiklik mullaseire keskkonnaseire andmete kogumise ja avalikustamise infosüsteemis KESE (National soil monitoring in environmental monitoring data collection and disclosure information system KESE	Y	-	Y	-	Y	Y	Y	Y	-	-	Y	Y	-	-	-	Y	-	Y	-	-	-	-	-	Y	-	Y	Y	Y	-	Y	-	Y	-	-	-	-	Y	Y	Y	Y	Y	-	-	-	-
	-	Regular monitoring of arable soils (I period)	Y	-	Y	-	Y	Y	Y	-	-	-	Y	-	-	-	-	Y	-	Y	-	-	-	-	-	Y	-	Y	Y	Y	-	Y	-	Y	-	-	-	-	-	Y	-	Y	-	-	-	-	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		SP1													SP2										SP3														SP4				SP5			
	Abbreviation	Name	SP1.1	SP1.2	SP1.3	SP1.4	SP1.5	SP1.6	SP1.7	SP1.8	SP1.9	SP1.10	SP1.11	SP1.12	SP1.13	SP2.1	SP2.2	SP2.3	SP2.4	SP2.5	SP2.6	SP2.7	SP2.8	SP2.9	SP2.10	SP3.1	SP3.2	SP3.3	SP3.4	SP3.5	SP3.6	SP3.7	SP3.8	SP3.9	SP3.10	SP3.11	SP3.12	SP3.13	SP3.14	SP4.1	SP4.2	SP4.3	SP4.4	SP5.1	SP5.2	SP5.3	SP5.4
Finland	*VALSE	National monitoring of arable soil chemical quality 1974-2018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	*VALSE subset	VALSE subset 2018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
France	RMQS	French Soil Monitoring Network (Réseau de Mesures de la Qualité des Sols en français)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	-	-	-	-	-	-	Y	-	-	Y	-	-	-	Y	-	Y	Y	Y	-	-	-	-	-	Y	Y	-	Y	-	-	Y	-
	IGCS	Soil Management and Conservation Inventory	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	BDAT	Soil test database	-	-	Y	Y	Y	Y	Y	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	BDGSF	Soil geographical database of France at 1:1,000,000	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-
	-	*GlobalSoilmap France pH*	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-
	-	*GlobalSoilmap France CEC*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	-	*GlobalSoilmap France depth*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	AWC BDGSF	Carte de la Réserve Utile en eau issue de la Base de Données Géographique des Sols de France	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	WRB BDGSF	Carte du nom de sol dominant en WRB issue de la Base de Données Géographique des Sols de France	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
	BDAT MAPS	Maps of the agronomic soil properties from BDAT	-	-	Y	Y	Y	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	RMQS MAPS ETM	Krigged maps of trace elements contents in French soil	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
	-	*GlobalSoilmap France organic carbon*	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
	-	*GlobalSoilmap France available water content*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	Y	-	-	-
Germany	BZE_LW	German Agricultural Soil Inventory	-	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
Hungary	SIMS	Hungarian Soil Information and Monitoring System	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	Y	Y	-	-	-	Y	Y	Y	-	-	-	-	Y	-	-	-	Y	-	-	-
	DOSoReMi	Digital, Optimized, Soil Related Maps and Information in Hungary (DOSoReMI.hu)	-	-	-	Y	Y	Y	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	MEPAR	Hungarian Identification System of Agricultural Parcels	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	Y	-
Ireland	Irish SIS	Irish Soil Information System	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	SQUARE	Soil Quality Assessment Research Project	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	-	Y	-	Y	-	Y	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	Tellus GSI	Tellus Geophysical and Geochemical Survey	-	-	-	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-
	NSDB	National Soil Database of Ireland	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
Italy	SISI-BADASUOLI	Italian Soil Information System	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	Y	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-	-	-	-
	-	Paper published data	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	Y	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-	-	-	-
	-	National surveys	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	Y	-	Y	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-	-	-	-
	-	Regional soil services	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	Y	-	Y	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-	-	-	-
	-	Private companies	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		SP1													SP2										SP3														SP4				SP5			
	Abbreviation	Name	SP1.1	SP1.2	SP1.3	SP1.4	SP1.5	SP1.6	SP1.7	SP1.8	SP1.9	SP1.10	SP1.11	SP1.12	SP1.13	SP2.1	SP2.2	SP2.3	SP2.4	SP2.5	SP2.6	SP2.7	SP2.8	SP2.9	SP2.10	SP3.1	SP3.2	SP3.3	SP3.4	SP3.5	SP3.6	SP3.7	SP3.8	SP3.9	SP3.10	SP3.11	SP3.12	SP3.13	SP3.14	SP4.1	SP4.2	SP4.3	SP4.4	SP5.1	SP5.2	SP5.3	SP5.4
Latvia	Digital profiles	Digital soil database (soil profiles)	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	Digital polygons	Digital soil database (soil polygons)	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	LUCAS	*LV LUCAS Topsoil 2009 data*	-	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	GeoChem	Geochemical Atlas of Latvia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	SPPS	Soil agrochemical research database of the State Information System for Monitoring of Agricultural Plants (SISMAP)	-	-	Y	-	Y	Y	Y	-	Y	-	-	-	Y	Y	-	Y	-	Y	Y	-	-	-	-	Y	Y	-	Y	-	-	Y	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-
	SPPS N	Database on the monitoring of mineral nitrogen in soils of the SISMAP	-	-	Y	-	-	-	-	-	Y	-	-	-	-	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	SILAVA	National forest inventory	-	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Pēterlauki	The long-term field trials "Pēterlauki"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-
	*Vecauce	The long-term field trials "Vecauce"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	AREI Priekuļi	The long-term field trials in State Priekuļi Plant Breeding Institute	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	AREI Stende	The long-term field trials in State Stende Cereals Breeding Institute	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	*ZMNĪ	Amelioration cadastre information system	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	LLU Skrīveri	The long-term field trials "Sidrabiņi"	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	LU ĢZZF	Universiy of Latvia fieldwork station	Y	-	Y	-	Y	Y	Y	-	Y	-	-	-	-	-	-	Y	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	Dārzkopības institūts	The long-term field trials in the Institute of Horticulture	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-	-
Lithuania	Dirv_DR10LT	Spatial Information Portal of Lithuania	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	-	Y	Y	-	-	Y	-	-	-	-	-	Y	Y	-	-	Y	Y	Y	Y
	Dirv_vert	Spatial Information Portal of Lithuania	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
	*Mel_DR10LT	Spatial Information Portal of Lithuania	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Netherlands	CC-NL	Carbon Change dataset of the Netherlands	Y	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	SL_LTE_Networks	-	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	Y	-
	BRO-BHR-p	Basisregistratie Ondergrond - Borehole research - profile descriptions	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	BRO-SFR	Basisregistratie Ondergrond - Soil face research	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	BRO-SGM	Basisregistratie Ondergrond - Soil map	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	PPS_BB - Bedrijvennetwerk	Publiek-Private Samenwerking Beter Bodembeheer - Bedrijvennetwerk	Y	-	Y	-	Y	Y	Y	-	-	-	Y	Y	-	-	-	-	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	PPS_BB - Bodemkwaliteit	Publiek-Private Samenwerking Beter Bodembeheer - Bodemkwaliteit	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	PPS_BB - iSQAPER	Publiek-Private Samenwerking Beter Bodembeheer - Interactive Soil Quality Assessment in Europe and China for Agricultural Productivity and Environmental Resilience	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	PPS_BB - MAK	Publiek-Private Samenwerking Beter Bodembeheer - Manure And Compost	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	Y	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	BOFEK	Bodemfysische eenheden kaart	-	-	-	-	Y	Y	Y	-	-	Y	-	Y	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	BIS - Soilhydrophysics	Bodemfysische datasets BIS	-	-	-	-	Y	Y	Y	-	-	Y	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
	-	Geochemische atlas van Nederland (Geo-chemical altas of the Netherlands)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Norway	-	Norwegian soil survey sample grid 9 x9 km on agricultural land	Y	Y	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	Norwegian soil survey	Y	Y	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	Norwegian soil property database	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Avalanche and rockslide - rough susceptiblitity zones	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-
	-	Flood hazard maps are prepared for floods with 20-, 200- and 1000-year return period	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		SP1													SP2										SP3														SP4				SP5			
	Abbreviation	Name	SP1.1	SP1.2	SP1.3	SP1.4	SP1.5	SP1.6	SP1.7	SP1.8	SP1.9	SP1.10	SP1.11	SP1.12	SP1.13	SP2.1	SP2.2	SP2.3	SP2.4	SP2.5	SP2.6	SP2.7	SP2.8	SP2.9	SP2.10	SP3.1	SP3.2	SP3.3	SP3.4	SP3.5	SP3.6	SP3.7	SP3.8	SP3.9	SP3.10	SP3.11	SP3.12	SP3.13	SP3.14	SP4.1	SP4.2	SP4.3	SP4.4	SP5.1	SP5.2	SP5.3	SP5.4
Poland	SAM25k	Soil Agricultural Map in scale 1:25000	-	-	-	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	ProfSAM25k	Referance soil profiles for SAM25	-	-	-	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Monit44k	Programme: "Assessment of ecological status of farmland in Poland"	-	-	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	Y	-	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	Monit41k		-	-	Y	Y	Y	Y	Y	Y	-	-	-	Y	-	Y	-	Y	-	Y	Y	-	-	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	-	-	-
	MChGO	Monitoring of the Chemical Properties of Arable Soils in Poland	-	-	Y	Y	Y	Y	Y	Y	Y	-	-	Y	Y	Y	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	-	-	-	-	-	Y	-	Y	-	-	-	-	-
	MonitFarm/Farm 500		-	-	Y	Y	Y	Y	Y	Y	-	Y	-	Y	-	Y	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	Y	-	-
Portugal	INFOSOLO	Database of Soil Profile Data for Portugal	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	Y	-	-	-
	PROPSOLO	Soil hydraulic properties of portuguese soils	Y	-	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-
	-	*ICP Forests level I plots in Portugal sampled for the BioSoil soil survey: Sampling at fixed depths; profile characterization*	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	Y	Y	-	Y	-	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-
Slovakia	BPEJ	Soil ecological units (land evaluation system)	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-	-	-	-	-	-	-	-	-	Y	-	Y	-	-	-	-	-
	CMS - P	National soil monitoring system	Y	-	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	Y	-	Y	-	-	-	-	-	Y	-	Y	Y	Y	-	Y	-	Y	Y	-	-	-	-	Y	Y	-	-	-	-	-	-
	PM400	Soil Map of Slovakiain 1:400k scale	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	AISOP	Agricultural Soil Profiles Database	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	*LPIS	National Land Parcel Identification System	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Slovenia	SOIL PROFILES SLO	Soil profiles of Slovenia	Y	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Soil fertility	Y	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	-	Y	-	Y	Y	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	-	Y	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Food, feed and fibre indicators	Y	-	Y	Y	Y	Y	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Erosion control indicators	Y	-	-	Y	Y	Y	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Climate regulation indicators	Y	-	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Indicators for labile organic carbon	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	Y	-
	CEBAS Tres Caminos	CSIC-LTE CEBAS Tres Caminos	-	-	Y	Y	Y	Y	Y	-	-	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
	CEBAS Cieza	CSIC-LTE CEBAS Cieza	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
	ICA La Poveda	CSIC-LTE ICA La Poveda	-	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
	Senés	CSIC-LTE EEAD Senes experimental farm	-	-	Y	-	-	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
	IRNASA	CSIC-LTE IRNASA Salamanca	Y	-	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	Y	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	Y	Y	-	Y	-	-	Y

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		SP1													SP2										SP3														SP4				SP5			
	Abbreviation	Name	SP1.1	SP1.2	SP1.3	SP1.4	SP1.5	SP1.6	SP1.7	SP1.8	SP1.9	SP1.10	SP1.11	SP1.12	SP1.13	SP2.1	SP2.2	SP2.3	SP2.4	SP2.5	SP2.6	SP2.7	SP2.8	SP2.9	SP2.10	SP3.1	SP3.2	SP3.3	SP3.4	SP3.5	SP3.6	SP3.7	SP3.8	SP3.9	SP3.10	SP3.11	SP3.12	SP3.13	SP3.14	SP4.1	SP4.2	SP4.3	SP4.4	SP5.1	SP5.2	SP5.3	SP5.4
Sweden	-	National Arable Soil & Crop Inventory	-	-	Y	Y	Y	Y	Y	-	Y	-	-	-	-	Y	-	-	-	Y	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	National Inventory of Arable Soils	-	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Soil Compaction	National Soil Compaction Survey	Y	-	Y	-	Y	Y	Y	-	-	-	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	-	Digital Soil Mapping Sweden	-	-	-	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Map of Soil Types	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Observationsfält på jordbruksmark (Monitoring water from arable fields)	-	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	Typområden på jordbruksmark (Monitoring water from agricultural catchments)	-	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	-	Jordlagerföljder (Soil layer data)	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Erosion risk maps	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Switzerland	-	Swiss Soil Dataset Version 5 (2020)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
Turkey	TAGEM_SFWRCRI_SIS	Turkish Ministry of Agriculture and Forestry Soil Fertilizer and Water Resources Central Research Institute Soil Information System	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	-	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	Y	-	Y	Y	-	-	-	-	Y	-	Y	-	-	-	Y	-
UK	NATMAP	National Soil Map of England and Wales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	NSI_Profile	National Soils Inventory - Profile	Y	-	-	-	Y	Y	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NSI_Topsoil	National Soils Inventory - Topsoil	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	CEH_Topsoil	Countryside survey of topsoil in Great Britain	-	-	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NSI_Scotland_1978-88	National Soil Inventory of Scotland (1978-88)	Y	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	Y	-	Y	Y	-	-	-	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	NSI_Scotland_2007-09	National Soil Inventory of Scotland (2007-09)	Y	-	Y	Y	Y	Y	Y	-	-	Y	-	Y	-	-	-	Y	-	Y	Y	-	-	-	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	NSM_Scotland	National soil map of Scotland 1:250000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
	NSM_Scotland_Partial	Soil map of Scotland (partial cover) 1:25000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	NSM_Northern_Ireland	AFBI Soil Series Map of Northern Ireland	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	AFBI_5K_Attributes	AFBI 5k Soil Attribute Dataset	Y	-	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
	UK_SHPS	UK Soil and Herbage Pollutant Survey	-	-	Y	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	TELLUS_A	TELLUS Regional A Horizon Soils	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	TELLUS_B	TELLUS Regional Subsoil	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	TELLUS_XRF	TELLUS Regional A Horizon Soils-XRF	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ESDAC point data	LUCAS	LUCAS Topsoil data (Soil Modules)	-	-	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	Y	-	-	-	Y	Y	Y	Y	-	Y	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	SPADE-2	Soil Profile Analytical Database for Europe	Y	-	Y	Y	Y	Y	Y	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
	SPADE/M	Measured soil profiles	Y	-	Y	-	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
	SPADE-14	Soil Profile Analytical Database 14	Y	-	Y	Y	Y	Y	Y	-	Y	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-
ESDAC	*Datasets for Soil Threats (detailed: JRC122248 Techn.R.*)*		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y		Y	Y	-	Y	-	-	Y	Y	-	Y	Y	-	-	Y	Y	-	-	-	-
	Datasets for Soil Functions (detailed: JRC122248 Techn.R.)*		-	-	-	-	-	-	-	-	-	-	-	Y	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	Y

*Panagos, P., Jones, A., Van Liedekerke, M., Orgiazzi, A., Lugato, E. and Montanarella, L., JRC support to the European Joint Programme for soil (EJP SOIL), EUR 30450 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-25322-8, doi:10.2760/74273, JRC122248.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Annex 2/C The thematic content of the databases for soil management (MG), reported by the partners.

Abbreviations and indications:

MG1. Soil tillage, farming system:
- MG1.1 Conventional/organic/other farming
- MG1.2 Precision farming
- MG1.3 Tillage (conventional/ reduced/ strip-till/ no-tillage)

MG2. Crop and cropping system:
- MG2.1 Crop (data for main crop)
- MG2.2 Cropping system (e.g. monoculture, kind of rotation etc.)

MG3. Soil nutrient management and plant protection:
- MG3.1 Fertilization (mineral/organic/both)
- MG3.2 Microbiological preparations
- MG3.3 Pesticides
- MG3.4 Soil conditioners related to soil /plant health *(other than MG3.2)*

MG4. Water management and related protection:
- MG4.1 Irrigation
- MG4.2 Drainage
- MG4.3 Soil conditioners related to water protection

MG5. Measures to control soil erosion:
- MG5.1 Terraces (wall, bench, ridges or raised beds, others)
- MG5.2 Windbreak hedges
- MG5.3 Runoff water management systems (channels, etc.)
- MG5.4 Buffer strips
- MG5.5 In field erosion control measures (e.g. micro-dams between ridges, etc.)
- MG5.6 Small scale buffering infrastructure (retention ponds, dams, etc.)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG6. Permanent crops:
- MG6.1 Green manuring
- MG6.2 Cover crops
- MG6.3 Mulching

MG7. Other management practices:
- MG7.1 Amelioration (other than drainage MG4.2)
- MG7.2 Greenhouses
- MG7.3 Other

* indicates reported MG dataset without further details

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		MG1			MG2		MG3				MG4			MG5						MG6			MG7		
	Abbreviation	Name	MG1.1	MG1.2	MG1.3	MG2.1	MG2.2	MG3.1	MG3.2	MG3.3	MG3.4	MG4.1	MG4.2	MG4.3	MG5.1	MG5.2	MG5.3	MG5.4	MG5.5	MG5.6	MG6.1	MG6.2	MG6.3	MG7.1	MG7.2	MG7.3
Austria	BORIS	Soil Information System	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Österreichische Bodenschätzung (Austrian Soil Condition Survey)	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Invekos - Agricultural Data	Y	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	AGES	AGES farm data and long term research sites	-	-	Y	Y	Y	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	BAW	BAW specialized project data and long term research	Y	Y	Y	Y	Y	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	IfÖL	IfÖL - long term research site	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Belgium Flanders	-	Uitgevoerde gemeentelijke erosiebestrijdingswerken	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	-	-	-	-	-	-
	-	Bemestingsallocatie	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Landbouwgebruikspercelen	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	-	Groeicurve	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Belgium Wallonia	CNSW	Digital Map of Walloon Soils	Y	-	-	Y	Y	Y	-	Y	-	-	Y	-	-	-	Y	Y	Y	-	Y	-	-	-	-	-
	CARBIOSOL	Carbiosol map	Y	-	Y	-	-	Y	-	Y	-	-	-	-	-	-	Y	Y	Y	-	Y	-	-	-	-	-
	REQUASUD	REQUASUD database	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Czech Republic	KPP	Komplexní průzkum půd (Systematic Soil Survey)	-	-	-	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	LPIS	Land Registry LPIS (Land Parcel Identification System) - AZZP	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	*Register of contaminated areas	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	*Basal soil monitoring - basic	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Denmark	DSMDB	Danish Soil Monitoring Data Base	-	-	Y	Y	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	DDSM	Danish Digital Soil Maps	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	DFR	Danish Farmers' Registrations	Y	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Estonia	KESE	Riiklik mullaseire keskkonnaseire andmete kogumise ja avalikustamise infosüsteemis KESE (National soil monitoring in environmental monitoring data collection and disclosure information system KESE	Y	-	Y	Y	Y	Y	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Regular monitoring of arable soils (I period)	Y	-	Y	-	Y	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	ARIB_register	The Agricultural Registers and Information Boards register of land use structure by land parcels	Y	-	-	Y	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		MG1			MG2		MG3				MG4			MG5						MG6			MG7		
	Abbreviation	Name	MG1.1	MG1.2	MG1.3	MG2.1	MG2.2	MG3.1	MG3.2	MG3.3	MG3.4	MG4.1	MG4.2	MG4.3	MG5.1	MG5.2	MG5.3	MG5.4	MG5.5	MG5.6	MG6.1	MG6.2	MG6.3	MG7.1	MG7.2	MG7.3
Finland	*VALSE	National monitoring of arable soil chemical quality 1974-2018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-
	*VALSE subset	VALSE subset 2018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-
France	RMQS	French Soil Monitoring Network (Réseau de Mesures de la Qualité des Sols en français)	Y	-	Y	Y	Y	Y	Y	Y	-	Y	Y	-	Y	Y	Y	Y	-	-	Y	Y	-	-		-
	BDGSF	Soil geographical database of France at 1:1,000,000	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-		-
Germany	BZE_LW	German Agricultural Soil Inventory	Y	-	Y	Y	Y	Y	-	-	-	Y	Y	-	Y	-	Y	-	-	-	Y	Y	Y	Y		-
	DESTATIS	Landwirtschaftszählung/ Agrarstrukturerhebung	Y	-	Y	Y	Y	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	Y	Y	-	-	Y	-
Hungary	-	Nitrate Database	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	MEPAR	Hungarian Identification System of Agricultural Parcels	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	WEB_GN	Farmer Diary Program on the Web (Webes Gazdálkodási Napló Program)	Y	Y	Y	Y	-	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	Y	Y	Y	-	-	-
Italy	RICA	Rete di Informazione Contabile Agricola (Farms Agronomical Mangement database)	Y	-	-	Y	-	Y	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		MG1			MG2		MG3				MG4			MG5						MG6			MG7		
	Abbreviation	Name	MG1.1	MG1.2	MG1.3	MG2.1	MG2.2	MG3.1	MG3.2	MG3.3	MG3.4	MG4.1	MG4.2	MG4.3	MG5.1	MG5.2	MG5.3	MG5.4	MG5.5	MG5.6	MG6.1	MG6.2	MG6.3	MG7.1	MG7.2	MG7.3
Latvia	SPPS	Soil agrochemical research database of the State Information System for Monitoring of Agricultural Plants (SISMAP)	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	SPPS N	Database on the monitoring of mineral nitrogen in soils of the SISMAP	-	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Pēterlauki	The long-term field trials “Pēterlauki”	Y	-	Y	Y	Y	Y	-	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	Vecauce	The long-term field trials “Vecauce”	Y	-	Y	-	Y	Y	-	Y	-	-	Y	-	-	-	-	Y	-	-	Y	-	Y	-	-	-
	AREI Priekuļi	The long-term field trials in State Priekuļi Plant Breeding Institute	Y	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-
	AREI Stende	The long-term field trials in State Stende Cereals Breeding Institute	-	-	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	ZMNĪ	Amelioration cadastre information system	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-
	LLU Skrīveri	The long-term field trials “Sidrabiņi”	Y	-	-	Y	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-
	Dārzkopības institūts	The long-term field trials in the Institute of Horticulture	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	-	-	Y	-	-	-	-	-	Y	Y	-	Y	-
Lithuania	Dirv_DR10LT	Spatial Information Portal of Lithuania	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	Y	-	-	Y	-	-	-	Y	Y	-
	*Dirv_vert	Spatial Information Portal of Lithuania	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Mel_DR10LT	Spatial Information Portal of Lithuania	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	Y	-	-	-	-	-	-	Y	-	-
Netherlands	SL_LTE_Networks	-	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-
	PPS_BB - Bodemkwaliteit	Publiek-Private Samenwerking Beter Bodembeheer - Bodemkwaliteit	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	PPS_BB - iSQAPER	Publiek-Private Samenwerking Beter Bodembeheer - Interactive Soil Quality Assessment in Europe and China for Agricultural Productivity and Environmental Resilience	-	-	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	PPS_BB - MAK	Publiek-Private Samenwerking Beter Bodembeheer - Manure And Compost	-	-	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-
Norway	AR5	Land resource map 1:5.000	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Direct Payments	Incentive Programs related to land use and animal husbandry	Y	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		MG1			MG2		MG3				MG4			MG5						MG6			MG7		
	Abbreviation	Name	MG1.1	MG1.2	MG1.3	MG2.1	MG2.2	MG3.1	MG3.2	MG3.3	MG3.4	MG4.1	MG4.2	MG4.3	MG5.1	MG5.2	MG5.3	MG5.4	MG5.5	MG5.6	MG6.1	MG6.2	MG6.3	MG7.1	MG7.2	MG7.3
Portugal	-	*ICP Forests level I plots in Portugal sampled for the BioSoil soil survey: Sampling at fixed depths; profile characterization*	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	COS 2018, 2018 Land Use and Soil Occupation Map for Portugal	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	Public irrigated areas 2018 and SIR (Irrigation Information System)	-	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Slovakia	BPEJ	Soil ecological units (land evaluation system)	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	CMS - P	National soil monitoring system	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	LPIS	National Land Parcel Identification System	Y	-	-	Y	Y	-	-	-	-	-	-	-	Y	Y	-	Y	-	-	-	Y	-	-	-	-
Slovenia	LPIS	LPIS-parcels in Slovenia	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
	MELIORATION SYS	Melioration systems of Slovenia	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Spain	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Soil fertility	Y	-	Y	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Food, feed and fibre indicators	Y	-	Y	Y	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Erosion control indicators	Y	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Climate regulation indicators	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	INIA-LTE La Canaleja	INIA-Experimental Farm La Canaleja (Madrid, Spain)- Indicators for labile organic carbon	Y	-	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	CEBAS Tres Caminos	CSIC-LTE CEBAS Tres Caminos	Y	-	Y	-	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	CEBAS Cieza	CSIC-LTE CEBAS Cieza	Y	-	Y	-	-	Y	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	ICA La Poveda	CSIC-LTE ICA La Poveda	Y	-	Y	-	Y	Y	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Senés	CSIC-LTE EEAD Senes experimental farm	Y	-	Y	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	IRNASA	CSIC-LTE IRNASA Salamanca	Y	-	Y	-	Y	Y	-	-	Y	Y	-	-	-	-	-	-	-	-	Y	-	-	-	-	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Country	Database		MG1			MG2		MG3				MG4			MG5						MG6			MG7		
	Abbreviation	Name	MG1.1	MG1.2	MG1.3	MG2.1	MG2.2	MG3.1	MG3.2	MG3.3	MG3.4	MG4.1	MG4.2	MG4.3	MG5.1	MG5.2	MG5.3	MG5.4	MG5.5	MG5.6	MG6.1	MG6.2	MG6.3	MG7.1	MG7.2	MG7.3
Sweden	Soil Compaction	National Soil Compaction Survey	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-
Sweden	-	Observationsfält på jordbruksmark (Monitoring water from arable fields)	Y	-	Y	Y	Y	Y	-	-	-	-	Y	-	-	-	-	Y	-	-	-	Y	-	-	-	-
Sweden	-	Typområden på jordbruksmark (Monitoring water from agricultural catchments)	Y	-	Y	Y	Y	Y	-	Y	-	-	Y	-	-	-	Y	Y	-	-	Y	Y	-	Y	-	-
Sweden	-	Skiftesdata (Arable field data)	Y	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	Y	-	-	Y	Y	-	-	-	-
UK	CROME	The Crop Map of England	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	CEH_LC+Fert	CEH Land Cover plus Fertilisers	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	CEH_LC+Pest	CEH Land Cover plus Pesticides	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	CEH_LC+Crops	CEH Land Cover plus Crops	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	England_Water_Situation	Water situation: national monthly reports for England	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	NVZ_England	Nitrate Vulnerable Zone designations 2017 to 2020	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	NVZ_Wales	Nitrate Vulnerable Zone designations 2017 to 2020	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	GB_Bare_Soil_Erosion	Bare Soil Water Erosion Potential	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	CEHLCM	CEH Land Cover Map 2017	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	England_Flood_Risk	Indicative Flood Risk Areas for England	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	NVZ_Scotland	Nitrate Vulnerable Zone designations 2017 to 2020	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	AWC_Scotland	Available Water Capacity Scotland	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	Erosion_Risk_Scotland	Map of soil erosion risk in Scotland (partial cover)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	SCOPES	Scottish Pesticide Surveys Database	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	Runoff_Risk_Scotland	Map of runoff risk in Scotland (partial cover)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	Compaction_Risk	Maps of Soil compaction risk in Scotland (partial cover)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
UK	NI_Agri_Census	Northern Ireland Agricultural Census	Y	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	AFBI_Pesticide_Usage	AFBI Pesticide usage monitoring surveys	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UK	Northern Ireland	The full area of Northern Ireland is identified as a NVZ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Annex 3 Soil property data

SP1.1-2. Profile depths, data availability in participating countries

Effective (plant exploitable) depth is the lower limit of biologic activity, which generally coincides with the common rooting depth.

SP 1.1-2 Profile depths	databases					method		method
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**total profile depth**	**described in field**	**plant exploitable depth**	**described in field**
Austria	5	FPO	eBOD	cm	5	yes	0	-
Belgium Flanders	2	F	DOV, SMF	cm	2	yes	0	-
Belgium Wallonia	2	P	-	cm	2	yes	0	-
Czech Republic	1	P	-	cm	1	yes	0	-
Denmark	1	P	-	cm	1	yes	0	-
Estonia	2	FR	KESE, SMI	cm	2	yes	0	-
Finland	2	P	-	cm	2	yes	0	-
France	4	FP	RMQS, BDGSF	cm, class	4	yes	3	yes
Germany	1	F	BZE_LW	cm	1	yes	1	yes
Hungary	1	P	-	cm	1	yes	1	yes
Ireland	2	PO	-	cm, m	2	yes	2	yes
Italy	5	FP	SISI, PPD, NS	cm	5	yes	5	yes
Latvia	3	FP	DigProf	cm	2	yes	1	yes
Lithuania	1	F	DR10LT	cm	1	yes	1	yes
Netherlands	3	O	-	cm	3	yes	1	yes
Norway	3	FP	NSS	cm	3	yes	3	yes
Poland	No data	-	-	-	0	-	0	-
Portugal	4	FP	INFSOL, PROSOL	cm	4	yes	1	yes
Slovakia	2	FP	BPEJ	class	2	yes	0	-
Slovenia	1	F	SPSLO	cm	1	yes	0	-
Spain	5	P	-	cm	5	yes	0	-
Sweden	2	FP	SOILCOM	cm, m	2	yes	0	-
Switzerland	1	F	SWISOIL	cm, class	1	yes	1	yes
Turkey	1	P	-	cm	1	yes	1	yes
United Kingdom	4	FRP	NSISC99, NSISC09, AFBI 5K, NSI_Prof	cm	4	yes	0	-
%*						92		48

Data policy : F-freely ; R- freely for research purpose EJP SOIL ; P-permission; O- other

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.3. Organic carbon content, data availability and applied determination methods reported by participating countries.

Soil organic carbon refers to the carbon component of organic compounds.

SP 1.3 Organic C	databases			applied method					
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**WET_WB**	**WET_TYURIN**	**WET_OTHER**	**DRY_W_LOSS**	**DRY_ADC**	**other**
Austria	4	FPO	eBOD	yes	no	yes	no	yes	no
Belgium Flanders	2	F	DOV, SOCMB	yes	no	no	no	yes	no
Belgium Wallonia	3	P	-	yes	no	no	yes	yes	no
Czech Republic	3	PO	-	yes	no	yes	no	yes	no
Denmark	5	RP	DDSM	no	no	no	yes	yes	no
Estonia	3	FRO	KESE, SMI	no	yes	no	no	yes	no
Finland	2	P	-	no	no	no	no	yes	no
France	4	FP	RMQS, BDAT	no	no	yes	no	yes	no
Germany	1	F	BZE_LW	no	no	no	no	yes	no
Hungary	1	P	-	no	yes	no	no	no	no
Ireland	3	PO	-	yes	no	no	no	yes	no
Italy	5	FP	SISI, PPD, NS	yes	no	yes	no	yes	no
Latvia	7	RP	LLU	no	yes	no	no	yes	no
Lithuania	1	F	DR10LT	no	yes	no	no	no	no
Netherlands	3	O	-	no	no	no	no	yes	yes
Norway	3	FP	NSS	no	no	no	no	yes	yes
Poland	4	FRP	MChGO, MonFrm	no	yes	no	no	yes	no
Portugal	4	FP	INFSOL, PROSOL	yes	no	yes	no	yes	no
Slovakia	2	P	-	yes	yes	no	no	yes	no
Slovenia	1	F	SPSLO	yes	no	no	no	no	no
Spain	5	P	-	yes	no	no	no	yes	no
Sweden	5	FP	SOILCOM	no	no	no	yes	yes	no
Switzerland	1	F	SWISOIL	yes	no	no	no	yes	no
Turkey	1	P	-	yes	no	yes	no	yes	no
United Kingdom	7	FRP	NSI_Top, NSISC88, NSISC09, AFBI 5K, TEL_XRF	yes	no	no	yes	yes	no
%*				52	24	24	16	88	8

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

applied method: **WET_WB**: wet combustion: Walkley Black (titrimetric); **WET_TYURIN**: wet combustion: Tyurin (titrimetric); **WET_OTHER**: wet combustion: other; **DRY_W_LOSS**: dry combustion, weight loss on ignition. **DRY_ADC**: automated dry combustion (CNS, TOC, Elemental Analyis (EA)). **other**: another method

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

SP1.4. PH in water, data availability and applied determination methods reported by participating countries

The pH (water) figure refers to the acidity of the soil solution, the pH scale is from 0 (most acid) to 14 (most alkaline) and a pH of 7 is neutral.

SP1.4 pH(water)	databases			applied method			
Country	**Relevant for topic**	**data policy**	**(at least a part of it) open access or freely available for EJP SOIL**	**1:2.5 SLR**	**1:5 SLR**	**SAT_EXTR**	**other**
Austria	4	FPO	eBOD	yes	yes	yes	no
Belgium Flanders	1	F	DOV	yes	yes	no	no
Belgium Wallonia	2	P	-	no	yes	no	no
Czech Republic	2	P	-	yes	yes	no	no
Denmark	3	RP	DDSM	yes	no	yes	no
Estonia	No data	-	-	-	-	-	-
Finland	1	P	-	yes	no	no	no
France	5	FP	RMQS, BDAT	no	yes	no	no
Germany	1	F	BZE_LW	no	yes	no	no
Hungary	2	P	-	yes	no	no	no
Ireland	2	PO	-	yes	no	no	no
Italy	5	FP	SISI, PPD, NS	yes	yes	yes	no
Latvia	2	P	-	no	yes	no	no
Lithuania	1	F	DR10LT	yes	yes	yes	no
Netherlands	3	O	-	no	no	no	
Norway	1	P	-	not specified			no
Poland	4	FRP	MChGO,MonFrm	yes	no	no	no
Portugal	4	FP	INFSOL, PROSOL	yes	yes	no	no
Slovakia	2	P	-	yes	yes	no	no
Slovenia	1	F	SPSLO	yes	yes		no
Spain	7	P	-	yes	yes	no	no
Sweden	4	P	-	no	yes	no	no
Switzerland	1	F	SWISOIL	not specified			no
Turkey	1	P	-	yes	yes	yes	no
United Kingdom	7	FRP	NSI_Top, NSISC88, NSISC09,AFBI 5K, UKSHPS, TEL_XRF	yes	yes	no	yes
%*				64	64	20	8

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **1:2.5 SLR**: 1:2.5 soil:liquid ratio (water); **1:5 SLR**: 1:5 soil:liquid ratio (water); **SAT_EXTR**: from saturated extract; **other**: another method.

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.5-7. Particle-size distribution, data availability and applied determination methods reported by participating countries.

Particle size distribution refers to the proportions by dry mass of a soil distributed over specified particle-size ranges.

SP1.5-7 PSD	databases			size			applied method								
Country	Relevant for topic	data policy	for EJP or freely	sand (mm)	silt (mm)	clay (mm)	sieve-pipette method	PIPETTE_NO_PRE:	PIPETTE_OM_CACO3_IRON	PIPETTE_OM_CACO3	PIPETTE_OM	hydro-meter method	HYDROM_OM_CACO3	LDM	other
Austria	5	FPO	eBOD	0.063-2	0.002-0.063	<0.002	yes	no	no	no	no	no	no	no	yes
Belgium Flanders	1	F	DOV	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	yes	no
Belgium Wallonia	2	P	-	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	no	yes
Czech Republic	4	PO	-	0.05-2	0.001-0.05	<0.001	yes	no	no	no	no	no	no	no	yes
Denmark	3	RP	DDSM	0.02-0.063, 0.063-0.125, 0.125-0.2, 0.2-2	0.002-0.02	<0.002	no	no	yes	no	no	no	no	no	no
Estonia	2	FR	KESE, SMI	0.063-2	0.002-0.063	<0.002	no	no	no	no	yes	no	no	no	no
Finland	2	P	-	0.06-2	0.002-0.06	<0.002	not specified								
France	3	FP	RMQS, BDAT	0.05-2	0.002-0.05	<0.002	no	yes	no	no	no	no	no	no	no
Germany	1	F	BZE_LW	0.063-2	0.002-0.063	<0.002	no	no	no		no	no	no	no	no
Hungary	2	P	-	0.05-2	0.002-0.05	<0.002	no	yes	no	no	no	no	no	no	no
Ireland	2	PO	-	0.05-2	0.002-0.05	<0.002	yes	no	no	yes	no	no	no	no	no
Italy	5	FP	SISI, PPD, NS	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Latvia	7	RPO	DSB	0.063-2	0.002-0.063	<0.002	yes	no	no	yes	no	no	no	yes	no
Lithuania	1	F	DR10LT	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Netherlands	5	FO	BOFEK, BIS-SH	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	yes
Norway	3	FO	NSS	0.06-2	0.002-0.06	<0.002	yes	no	no	no	no	no	no	no	yes
Poland	6	FRP	MChGO, MonFrm	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	yes	yes
Portugal	4	FP	INFSOL, PROSOL	0.02-2	0.002-0.02	<0.002	yes	no	yes	yes	no	no	no	no	no
Slovakia	2	P	-	0.05-2	0.01-0.05	<0.01	yes	no	no	no	no	no	no	no	no
Slovenia	1	F	SPSLO	0.05-0.2, 0.2-2	0.002-0.02, 0.02-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Spain	8	P	-	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	no	no
Sweden	6	FP	SOILCOM	0.06-2	0.002-0.06	<0.002	no	no	no		no	no	no	no	yes
Switzerland	1	F	SWISOIL	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	no	no	no	no
Turkey	1	P	-	0.05-2	0.002-0.05	<0.002	no	no	no	no	no	no	yes	no	no
United Kingdom	4	FRP	NSISC88, NSISC09, AFBI 5K,	0.05-2	0.002-0.05	<0.002	yes	no	no	no	no	yes	no	no	no
%*							68	8	8	20	4	16	4	12	28

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **sieve-pipette method**: sieve-pipette method, no details; **PIPETTE_NO_PRE**: sieve-pipette method no pretreatment; **PIPETTE_OM_CACO3_IRON**: sieve-pipette method OM, CaCO3 and iron oxides removal; **PIPETTE_OM_CACO3:** sieve-pipette method OM and CaCO3 removal; **PIPETTE_OM:** sieve-pipette method OM removal; **hydrometer method**: hydrometer method, no details; **HYDROM_OM_CACO3**: hydrometer method, OM and CaCO3 removal; **LDM**: laser diffraction method), no details; **other:** another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.8. Gravel content, data availability and applied determination methods reported by participating countries

It refers to the proportion by mass or volume of coarse fragments distributed in soil, fraction of the soil material > 2 mm.

SP1.8 Gravel	databases					applied method	
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**description**	**size/unit**	**mass based measurement**	**other**
Austria	2	P	-	shape/ size description	vol%, mm	yes	yes
Belgium Flanders	1	F	DOV	-	>2 mm	yes	no
Belgium Wallonia	2	P	-	-	%	no	yes
Czech Republic	1	P	-	categories	-	no	yes
Denmark	1	P	-	-	% dry wgt	yes	no
Estonia	2	F	SM, KESE	classes	m3 ha-1	no	yes
Finland	2	P	-	-	>2 mm	yes	no
France	3	FP	RMQS, BDGSF	phase	g 100g-1, %	yes	yes
Germany	1	F	BZE_LW	-	vol%	yes	yes
Hungary	1	P	-	-	m3 m-3	yes	no
Ireland	2	PO	-	-	m3 m-3	yes	no
Italy	5	FP	SISI, PPD, NS	-	m3 m-3	yes	no
Latvia	No data	-	-	-	-	-	-
Lithuania	No data	-	-	-	-	-	-
Netherlands	No data	-	-	-	-	-	-
Norway	3	FP	NSS	-	%	yes	yes
Poland	4	FRP	MChGO, MonFrm	-	>1 mm, >2mm	yes	no
Portugal	4	FP	INFSOL, PROSOL	-	%	yes	no
Slovakia	2	P	-	-	%	no	yes
Slovenia	1	F	SPSLO	-	>2 mm	yes	no
Spain	No data	-	-	-	-	-	-
Sweden	1	P	-	.	2 - 20 mm, %	yes	no
Switzerland	1	F	SWISOIL	-	vol%	yes	no
Turkey	No data	-	-	-	-	-	-
United Kingdom	2	RP	AFBI 5K	-	vol%, %, >2 mm	no	yes
%*						60	36

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.9. Effective cation exchange capacity, data availability and applied determination methods reported by participating countries

The effective cation exchange capacity (ECEC) is defined as the total amount of exchangeable cations, which are mostly Na, K, Ca and Mg (termed as bases) in non-acidic soils and bases plus Al in acidic soils.

SP1.9 ECEC	databases			applied method			
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**BARIUM_CL**	**AMMONIUM_AC**	**other**
Austria	3	PO	-	mmol IÄ kg-1	yes	yes	no
Belgium Flanders	No data	-	-	-	-	-	-
Belgium Wallonia	1	P	-	cmol(+)kg-1	yes	no	no
Czech Republic	3	PO	-	cmolc kg-1, mmol ch. Ekv kg-1, mekv kg-1	yes	no	yes
Denmark	1	P	-	mol 100 g-1	no	yes	yes
Estonia	No data	-	-	-	-	-	-
Finland	No data	-	-	-	-	-	-
France	5	FP	RMQS, BDAT	cmolc kg-1, meq 100g-1	no	yes	yes
Germany	No data	-	-	-	-	-	-
Hungary	1	P	-	cmolc kg-1	yes	yes	no
Ireland	2	PO	-	cmolc kg-1	yes	yes	no
Italy	5	FP	SISI, PPD, NS	cmolc kg-1	yes	yes	no
Latvia	4	P	-	cmol(p+) kg-1, mval 100g-1	yes	no	no
Lithuania	No data	-	-	-	-	-	-
Netherlands	2	O	-	mmol+ kg-1	no	no	yes
Norway	1	P	-	cmolc kg-1	not specified		
Poland	2	FP	MChGO	mmol(+) 100g-1	no	yes	yes
Portugal	4	FP	INFSOL, PROSOL	cmolc kg-1, meq 100g-1	yes	yes	no
Slovakia	2	P	-	cmolc kg-1, mval 100 g-1	yes	no	yes
Slovenia	1	F	SPSLO	mmol 100g-1	yes	no	no
Spain	2	P	-	cmolc kg-1	yes	yes	no
Sweden	1	P	.	cmolc kg-1	yes	no	no
Switzerland	1	F	SWISOIL	mmolckg-1	yes	no	yes
Turkey	1	P	-	cmolc kg-1	yes	yes	no
United Kingdom	1	R	AFBI 5K	meq 100g-1	no	yes	no
%*					56	44	28

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **BARIUM_CL**: Cations extracted using Barium Chloride (BaCl2) plus exchangeable H + Al; **AMMONIUM_AC**: Cations extracted using Ammonium Acetate plus exchangeable H + Al; ; **other:** another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

SP1.10. Bulk density, data availability and applied determination methods reported by participating countries

Bulk density is calculated as the dry weight of soil divided by its volume, it can act as a basic indicator of soil compaction.

SP1.10-11 Bulk density	databases						bulk density of the fine earth				bulk density of the whole soil		
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**bulk density of the fine earth**	**bulk density of the whole soil in situ**	**unit**	**ARTIFIC_CORE**	**GRAVEL_CORE**	**EST_**	**calculated**	**UNDIST_CORE:**	**EST_**	**other**
Austria	3	P	-	3	0	dB g cm-3, %	not specified				-	-	-
Belgium Flanders	No data	-	-	0	0	-	-	-	-	-	-	-	-
Belgium Wallonia	No data	-	-	0	0	-	-	-	-	-	-	-	-
Czech Republic	1	P	-	0	1	g cm-3	-	-	-	-	yes	no	no
Denmark	3	RP	DDSM	0	3	g cm-3	-	-	-	-	yes	no	no
Estonia	2	FR	KESE, SMI	0	2	g cm-3	-	-	-	-	no	no	yes
Finland	2	P	-	1	1	g cm-3	not specified				yes	no	no
France	2	FP	RMQS	2	1	g cm-3	no	yes	no	yes	yes	no	yes
Germany	1	F	BZE_LW	1	1	g cm-3	no	no	no	yes	yes	no	no
Hungary	1	P	-	0	1	Mg m-3	-	-	-	-	yes	no	no
Ireland	2	PO	-	0	2	Mg m-3	-	-	-	-	yes	no	no
Italy	5	FP	SISI, PPD, NS	0	5	Mg m-3	-	-	-	-	yes	no	no
Latvia	No data	-	-	0	0	-	-	-	-	-	-	-	-
Lithuania	1	F	DR10LT	0	1	Mg m-3	-	-	-	-	yes	no	no
Netherlands	6	FO	BOFEK	2	4	g cm-3, kg m-3	not specified				yes	no	no
Norway	1	P	-	1	1	-	no	no	yes	no	no	yes	no
Poland	1	R	MonFrm	1	0	-	not specified				-	-	-
Portugal	4	FP	INFSOL, PROSOL	2	2	g cm-3, kg m-3	not specified		no	no	yes	no	no
Slovakia	1	P	-	0	1	g.cm-3	-	-	-	-	yes	no	no
Slovenia	No data	-	-	0	0	-	-	-	-	-	-	-	-
Spain	7	P	-	0	7	Mg m-3, g cm-3	-	-	-	-	yes	no	no
Sweden	1	F	SOILCOM	0	1	g cm-3	-	-	-	-	yes	no	no
Switzerland	1	F	SWISOIL	1	1	g cm-3	not specified				yes	no	no
Turkey	1	P	-	1	0	-	yes	no	no	no	-	-	-
United Kingdom	3	FP	NSISC09, UKSHPS	2	1	t m-3	yes	no	no	no	no	no	yes
%*							8	4	4	8	60	4	12

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **ARTIFIC_CORE**: disturbed/artificial soil core; **GRAVEL_CORE**: measured volume/absolut dry mass (at 105^{o}Celsius) of gravels g/cm3; **EST**: pedotransfer based estimation **calculated**: calculated from bulk density from whole soil and coarse elements mass; **UNDIST_CORE**: undisturbed core sample; **other:** another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.12. Available water capacity (plant available water), data availability and applied determination methods reported by participating countries

Available water capacity or plant available water capacity represents the volume of water stored within the soil available to the plant.

SP 1.12 AWC	databases				applied method			
Country	Relevant for topic	data policy	for EJP or freely	unit	UNDIST_CORE	EVAP_UNDIST_CORE	EST	other
Austria	4	FP	eBOD	%, mm, desc	no	no	yes	no
Belgium Flanders	No data	-	-	-	-	-	-	-
Belgium Wallonia	No data	-	-	-	-	-	-	-
Czech Republic	No data	-	-	-	-	-	-	-
Denmark	2	RP	DDSM	g 100 cm-3	yes	no	no	yes
Estonia	1	F	KESE	%	no	yes	no	no
Finland	0	-	-	-	-	-	-	-
France	1	F	AWC BDGSF	class	no	no	yes	no
Germany	1	F	BZE_LW	mm	no	no	yes	no
Hungary	1	P	-	-	no	no	yes	no
Ireland	No data	-	-	-	-	-	-	-
Italy	No data	-	-	-	-	-	-	-
Latvia	No data	-	-	-	-	-	-	-
Lithuania	1	F	DR10LT	%	yes	no	no	no
Netherlands	3	FO	BOFEK	g kg-1	no	yes	no	no
Norway	No data	-	-	-	-	-	-	-
Poland	3	FRP	MChGO, MonFrm	-	not specified			
Portugal	2	F	INFSOL, PROSOL	cm3 cm-3	yes	yes	yes	no
Slovakia	No data	-	-	-	-	-	-	-
Slovenia	No data	-	-	-	-	-	-	-
Spain	1	P	-	cm3 cm-3	no	no	yes	no
Sweden	No data	-	-	-	-	-	-	-
Switzerland	1	F	SWISOIL	class	not specified			
Turkey	No data	-	-	-	-	-	-	-
United Kingdom	1	F	NSISC09	mm	not specified			
%*					12	12	24	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **UNDIST_CORE**: undisturbed core samples at various suction (tension) values; **EVAP_UNDIST_CORE**: undisturbed core samples, continuous pF curve measured by evaporation method; **EST**: pedotransfer based estimation; **other**: another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP1.13. Electrical conductivity (EC), data availability and applied determination methods reported by participating countries

Electrical conductivity (EC) is a measure of the ability of the soil to conduct an electrical current, it indicates the amount of salts in soil (salinity of soil).

SP 1.13 EC	databases				applied method			
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**COND_SAT_PASTE**	**COND_1_2.5**	**COND_1_5**	**other**
Austria	3	PO	-	mS m-1	yes	yes	yes	no
Belgium Flanders	No data	-	-	-	-	-	-	-
Belgium Wallonia	No data	-	-	-	-	-	-	-
Czech Republic	No data	-	-	-	-	-	-	-
Denmark	No data	-	-	-	-	-	-	-
Estonia	No data	-	-	-	-	-	-	-
Finland	1	?	?	mS m-1	no	yes	no	no
France	1	F	RMQS	mS m-1	no	no	no	yes
Germany	1	F	BZE_LW	mS m-1	no	no	yes	no
Hungary	1	P	-	mS m-1	yes	yes	yes	no
Ireland	2	PO	-	mS m-1	yes	yes	yes	no
Italy	5	FP	SISI, PPD, NS	mS m-1	yes	yes	yes	no
Latvia	1	P	-	mS m-1	not specified			
Lithuania	1	F	DR10LT	mS m-1	not specified			
Netherlands	No data	-	-	-	-	-	-	-
Norway	No data	-	-	-	-	-	-	-
Poland	1	F	MChGO	mS m-1	no	no	yes	no
Portugal	2	F	INFSOL, PROSOL	mS cm-1	yes	yes	yes	no
Slovakia	1	P	-	mS m-1	yes	yes	yes	no
Slovenia	No data	-	-	-	-	-	-	-
Spain	1	P	-	dS m-1	no	no	yes	no
Sweden	No data	-	-	-	-	-	-	-
Switzerland	1	F	SWISOIL	mS m-1	no	no	yes	no
Turkey	1	P	-	mS m-1	yes	yes	yes	no
United Kingdom	No data	-	-	-	-	-	-	-
%*					28	32	44	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **COND_SAT_PASTE**: Measuring EC with a conductivity meter in a saturated paste of soil and water; **COND_1_2.5**: Measuring EC with a conductivity meter in a soil-water extract based on a 1:2.5 soil:liquid ratio; **COND_1_5**: Measuring EC with a conductivity meter in a soil-water extract based on 1:5 soil:liquid ratio; **other**: another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.1. Calcium carbonate content ($CaCO_3$), data availability and applied determination methods reported by participating countries

SP 2.1 $CaCO_3$	databases				applied method			
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**TITRATION**	**VOLUMETRIC**	**TGA**	**other**
Austria	5	FPO	eBOD	%	no	yes	no	no
Belgium Flanders	1	F	DOV	%	not specified			
Belgium Wallonia	1	P	-	mg kg-1	yes	no	no	no
Czech Republic	2	PO	-	%	yes	yes	no	no
Denmark	3	P	-	%	yes	no	no	no
Estonia	No data	-	-	-	-	-	-	-
Finland	No data	-	-		-	-	-	-
France	3	FP	RMQS, BDAT	g kg-1	no	yes	no	no
Germany	1	F	BZE_LW	g kg-1	no	no	yes	no
Hungary	1	P	-	g kg-1	no	yes	no	no
Ireland	2	PO	-	effervescence	no	no	no	yes
Italy	5	FP	SISI, PPD, NS	g kg-1	no	yes	no	no
Latvia	7	RPO	Dig Prof, LLU	presence, g kg-1, %	no	yes	no	yes
Lithuania	1	F	DR10LT	g kg-1	no	yes	no	no
Netherlands	No data	-	-	-	-	-	-	-
Norway	No data	-	-	-	-	-	-	-
Poland	4	FRP	MChGO, MonFrm	%	no	yes	no	no
Portugal	3	FP	INFSOL, PROSOL	g kg-1, %	no	yes	no	no
Slovakia	2	P	-	g kg-1	no	yes	no	no
Slovenia	No data	-	-	-	-	-	-	-
Spain	1	P	-	g kg-1	no	yes	no	no
Sweden	3	P	-	%	no	no	no	yes
Switzerland	1	F	SWISOIL	%	yes	yes	no	yes
Turkey	1	P	-	g kg-1	no	yes	no	no
United Kingdom	1	P	-	%	no	no	no	yes
%*					16	52	4	20

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **TITRATION**: Rapid titration method (Piper); **VOLUMETRIC**: Volumetric method (Scheibler); **CALCIMETER**: Calcimeter method (Bernard); **TGA**: thermogravimetric analysis; **other**: another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.2. Field capacity, data availability and applied determination methods reported by participating countries

Field capacity is the amount of water which can be hold against gravity in the soil.

SP 2.2 Field capacity (mm)	databases				applied method			
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**SWP_33_PLATE**	**SWP_OTHER**	**EST**	**other**
Austria	4	FP	eBOD	mm	no	no	yes	no
Belgium Flanders	No data	-	-	-	-	-	-	-
Belgium Wallonia	No data	-	-	-	-	-	-	-
Czech Republic	No data	-	-	-	-	-	-	-
Denmark	1	P	-	g 100 cm-3	no	yes	no	no
Estonia	No data	-	-	-	-	-	-	-
Finland	No data	-	-	-	-	-	-	-
France	No data	-	-	-	-	-	-	-
Germany	1	F	BZE_LW	mm	no	no	yes	no
Hungary	1	P	-	mm	yes	no	no	no
Ireland	No data	-	-	-	-	-	-	-
Italy	No data	-	-	-	-	-	-	-
Latvia	No data	-	-	-	-	-	-	-
Lithuania	1	F	DR10LT	mm	no	no	no	yes
Netherlands	2	FO	BOFEK	-	not specified			
Norway	No data	-	-	-	-	-	-	-
Poland	No data	-	-	-	-	-	-	-
Portugal	2	F	INFSOL, PROSOL	mm	no	no	yes	no
Slovakia	No data	-	-	-	-	-	-	-
Slovenia	No data	-	-	-	-	-	-	-
Spain	4	P	-	% vol, g kg-1, cm3 cm-3	no	no	yes	yes
Sweden	1	F	SOILCOM	1 m	not specified			
Switzerland	No data	-	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-	-
United Kingdom	No data	-	-	-	-	-	-	-
%*					4	4	16	8

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **SWP_33_PLATE**: soil water potential of -33 kPa, with ceramic plate; **SWP_OTHER**: soil water potential other than -33 kPa; **EST**: pedotransfer based estimation; **other**: another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.3 PH KCl, data availability and applied determination methods reported by participating countries

The pH (KCl) figure refers to the acidity in the soil solution, plus the reserve acidity in the colloids. pH (KCl) is always more acid than pH (water).

SP2.3 pH(KCl)	databases					applied method		
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**in 1 M KCl solution**	**in 0,01 M CaCl2-solution**	**1:2.5 SLR**	**1:5 SLR**	**SAT_EXTR**
Austria	5	FPO	eBOD	-	5	not specified		
Belgium Flanders	1	F	DOV	1	-	yes	yes	no
Belgium Wallonia	2	P	-	2	-	no	yes	no
Czech Republic	4	PO	-	1 (0.2M KCl)	3	yes	no	no
Denmark	1	P	-	-	1	not specified		
Estonia	3	FRO	KESE, SMI	3	-	no	yes	no
Finland	No data	-	-	-	-	-	-	-
France	1	P	-	1	-	no	yes	no
Germany	No data	-	-	-	-	-	-	-
Hungary	1	P	-	1	-	yes	no	no
Ireland	1	F	Tellus GSI	-	1	no	no	yes
Italy	4	FP	SISI, PPD, NS	4	-	yes	yes	yes
Latvia	5	RPO	LLU	5	$BaCl_2$	no	yes	no
Lithuania	1	F	DR10LT	1	-	yes	no	no
Netherlands	3	O	-	1	2	not specified		
Norway	No data	-	-	-	-	-	-	-
Poland	4	FRP	MChGO, MonFrm	4	-	yes	no	no
Portugal	No data	-	-	-		-	-	-
Slovakia	2	P	-	2	-	yes	no	no
Slovenia	1	F	SPSLO	1	-	not specified		
Spain	No data	-	-	-	-	-	-	-
Sweden	No data	-	-	-		-	-	-
Switzerland	1	F	SWISOIL	-	-	not specified		
Turkey	No data	-	-	-	-	-	-	-
United Kingdom	2	F	NSISC88, NSISC09	-	2	1:3 SLR	no	no
%*						32	24	8

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **1:2.5 SLR**: 1:2.5 soil:liquid ratio (1M KCl); **1:5 SLR**: 1:5 soil:liquid ratio (1M KCl); **SAT_EXTR**: from saturated extract

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.4 Saturated hydraulic conductivity (KSat), data availability and applied determination methods reported by participating countries

Saturated hydraulic conductivity describes water movement through saturated porous media.

SP 2.4 Ksat	databases				applied method		
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**laboratory_ constant**	**In_situ_ small**	**other**
Austria	2	FP	eBOD	cm d-1	not specified		
Belgium Flanders	No data	-	-	-	-	-	-
Belgium Wallonia	No data	-	-	-	-	-	-
Czech Republic	No data	-	-	-	-	-	-
Denmark	No data	-	-	-	-	-	-
Estonia	No data	-	-	-	-	-	-
Finland	No data	-	-	-	-	-	-
France	No data	-	-	-	-	-	-
Germany	No data	-	-	-	-	-	-
Hungary	No data	-	-	-	-	-	-
Ireland	1	P	-	-	not specified		
Italy	No data	-	-	-	-	-	-
Latvia	No data	-	-	-	-	-	-
Lithuania	1	F	DR10LT	cm h-1	yes	no	no
Netherlands	2	FO	BOFEK	-	not specified		
Norway	No data	-	-	-	-	-	-
Poland	No data	-	-	-	-	-	-
Portugal	1	F	PROSOL	cm d-1	no	no	yes
Slovakia	No data	-	-	-	-	-	-
Slovenia							
Spain	1	P	-	mm h-1	no	yes	no
Sweden	1	F	SOILCOM	cm hr-1	not specified		
Switzerland	1	F	SWISOIL	m s-1	not specified		
Turkey	No data	-	-	-	-	-	-
United Kingdom	No data	-	-	-	-	-	-
%*					4	4	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **laboratory_constant**: laboratory_constant head; **In_situ_small**: In_situ_small scale methods (augerhole, piezometer, guelph, double ring, pumped borehole, infiltrometer, inversed augerhole)

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.5. Plant available nutrients, data availability and applied determination methods reported by participating countries

SP 2.5 Plant available nutrients	databases				applied method		
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**elements**	**CLASSIC**	**ICP**	**other**
Austria	2	PO	-	-	yes	no	no
Belgium Flanders	No data	-	-	-	-	-	-
Belgium Wallonia	1	P	-	N, P, K, Ca, Mg	yes	yes	yes
Czech Republic	3	PO	-	P, K, Ca, Mg, Cu, Mn, Zn, Fe, Cd, S, B	yes	no	yes
Denmark	3	P	-	N, P	yes	no	no
Estonia	3	FRO	KESE,	-	yes	no	no
Finland	1	P	-	-	yes	no	no
France	No data	-	-	-	-	-	-
Germany	No data	-	-	-	-	-	-
Hungary	No data	-	-	-	-	-	-
Ireland	No data	-	-	-	-	-	-
Italy	5	FP	SISI, PPD, NS	P, K	not specified		
Latvia	9	RPO	LLU	N, P, K, Ca, Mg, Cu, Mn, Zn, S, B, Na, Al, Fe	yes	no	no
Lithuania	1	F	DR10LT	-	yes	no	no
Netherlands	4	O	-	N, P, K, Ca, Mg, Fe, Al	yes	yes	yes
Norway	No data	-	-	-	-	-	-
Poland	3	FRP	MChGO, MonFrm	P, K, Mg, Zn, Cu, Mn, Fe, B	yes	no	no
Portugal	No data	-	-	-	-	-	-
Slovakia	2	P	-	P, K, Mg, Cu, Zn, Mn	yes	no	no
Slovenia	No data	-	-	-	-	-	-
Spain	2	P	-	P, K, Ca, Mg	yes	no	no
Sweden	3	P	-	P, K, Al, Fe, Ca, Mg, Na	no	yes	no
Switzerland	No data	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-
United Kingdom	5	FRP	NSI_Top, CEH_Top, NSISC88, NSISC09, AFBI 5K	N, P, K, Mg, Cu, Mn, Zn, Co, Pb, Ni, Na, Si, Se	yes	yes	yes
%*					52	16	16

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **CLASSIC:** classic chemical laboratory methods; **ICP**: inductively coupled plasma; **other**: another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.6. Total amounts of elements in soil, data availability and applied determination methods reported by participating countries

SP 2.6 Total amounts of elements	databases				applied method		
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**elements**	**CLASSIC**	**ICP**	**other**
Austria	2	PO	-	-	yes	no	no
Belgium Flanders	No data	-	-	-	-	-	-
Belgium Wallonia	1	P	-	Cu, Fe, Mn, Mo, Zn	yes	yes	no
Czech Republic	2	PO	-	N, P, K, C, Ca, Mg, Cu, Mn, Zn, Fe, S, B	yes	yes	yes
Denmark	4	P	-	N, P, K, Na, Ca, Mg	yes	no	no
Estonia	No data	-	-	-	-	-	-
Finland	No data	-	-	-	-	-	-
France	No data	-	-	-	-	-	-
Germany	No data	-	-	-	-	-	-
Hungary	No data	-	-	-	-	-	-
Ireland	3	PO	-	N, P, K, Ca, Mg, C, Na, S, Cu, Mn, Zn, Si, Se, B	yes	no	no
Italy	5	FP	SISI, PPD, NS	N	not specified		
Latvia	7	PO	-	N, C, Al, Fe, Mn	yes	no	no
Lithuania	1	F	DR10LT	P, K, Ca, Mg	no	no	yes
Netherlands	5	O	-	N, P, K, C , Ca, Mg	yes	yes	yes
Norway	No data	-	-	-	-	-	-
Poland	4	FRP	MChGO, MonFrm	Fe, Mn, Al, Cu, Ni, Cr, Zn, Cd, Co, Pb	no	yes	no
Portugal	4	FP	INFSOL, PROSOL	N, P, K, Ca, Mg, Na, Cd, Cr, Cu, Ni, Pb, Zn	yes	yes	yes
Slovakia	No data	-	-	-	-	-	-
Slovenia	No data	-	-	-	-	-	-
Spain	1	P	-	N	yes	no	no
Sweden	3	P	-	P, K, As, Cd, Co, Cr, Cs, Cu, Hg, Mn, Mo, Ni, Pb, Sr, V, Zn, Se, B	no	yes	no
Switzerland	1	F	SWISOIL	Na, Ca, B, P, Mg, K, N, S, Se, Al	yes	yes	yes
Turkey	1	P	-	N,P,K, Fe, Cu, Zn, Mn,Cd, Cr, Ni, Pb	no	yes	yes
United Kingdom	7	FRP	NSI_Top, NSISC88, NSISC09, AFBI 5K	P, K, Mg, Ca, Cu, Mn, Zn, Co, Pb, Ni, Si, Se	yes	yes	yes
%*					44	36	28

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **CLASSIC:** classic chemical laboratory methods; **ICP**: inductively coupled plasma; **other**: another method

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

SP2.7 Quality of clay minerals, data availability and applied determination methods reported by participating countries

SP 2.7. Quality of clay minerals	databases				method
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**unit**	**XRD**
Lithuania	1	F	DR10LT	%	yes
%*					4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **XRD**: X-ray diffractometry

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.8. Distribution of soil organisms, data availability and applied determination methods reported by participating countries

SP 2.8 Dist. of soil organisms	databases				
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**soil organism**	**applied method**
Austria	No data	-	-	-	-
Belgium Flanders	No data	-	-	-	-
Belgium Wallonia	No data	-	-	-	-
Czech Republic	No data	-	-	-	-
Denmark	No data	-	-	-	-
Estonia	No data	-	-	-	-
Finland	No data	-	-	-	-
France	1	F	RMQS	earthworms	estimated at field
Germany	No data	-	-	-	-
Hungary	No data	-	-	-	-
Ireland	2	PO	-	fungal, bacterial, microbial biomass, earthworms, soil enzymes	respiration activity, counting, sequencing
Italy	No data	-	-	-	-
Latvia	1	P	-	earthworms, soil microbiota	enzymatic activity
Lithuania	No data	-	-	-	-
Netherlands	3	O	-	nematodes, earthworms; microbial biomass carbon	counting: microscope and visual, fumigation, extraction
Norway	No data	-	-	-	-
Poland	No data	-	-	-	-
Portugal	No data	-	-	-	-
Slovakia	No data	-	-	-	-
Slovenia	No data	-	-	-	-
Spain	No data	-	-	-	-
Sweden	No data	-	-	-	-
Switzerland	1	F	SWISOIL	earthworms	manual counting
Turkey	No data	-	-	-	-
United Kingdom	1	P	-	soil invertebrtaes	not specified
%*				24	

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

The notation "not specified" indicates that there is a measurement on it, but the method is not reported.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.10. Properties for NIR and MIR (near and mid infrared), data availability reported by participating countries.

SP 2.10 Prop. for NIR and MIR	databases			
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**applied method**
Austria	No data	-	-	-
Belgium Flanders	No data	-	-	-
Belgium Wallonia	No data	-	-	-
Czech Republic	No data	-	-	-
Denmark	1	P	-	Hyperspectral soil reflectance
Estonia	No data	-	-	-
Finland	No data	-	-	-
France	1	F	RMQS	NIR and MIR (topsoil)
Germany	1	F	BZE_LW	NIR spectra are recorded (90% completed)
Hungary	No data	-	-	-
Ireland	No data	-	-	-
Italy	2	FP	NS	Lab measurements
Latvia	No data	-	-	-
Lithuania	No data	-	-	-
Netherlands	3	O	-	Lab measurements
Norway	No data	-	-	-
Poland	1	F	MChGO	Lab measurements
Portugal	No data	-	-	-
Slovakia	No data	-	-	-
Slovenia				
Spain	No data	-	-	-
Sweden	2	P	-	vis-NIR (400-2500 nm), PXRF
Switzerland	No data	-	-	-
Turkey	No data	-	-	-
United Kingdom	No data	-	-	-
%*				28

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

applied method: **NIR**: Near-infrared; **MIR**: mid-infrared; **vis-NIR**: visible-Near-infrared; **PXRF**: Portable X-ray fluorescence analyzer

The notation "not specified" indicates that there is a measurement on it, but the method is not reported

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP2.11-12. Other country specific data, their availability and other mentions reported by participating countries

SP 2.11-12 other or country-specific	databases			
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	mentions
Austria	1	P	-	1. contaminants (Heavy metals, organic pollutants)
Belgium Flanders	2	F	DOV, SMF	1. soil texture class
Belgium Wallonia	2	P	-	1. C/N ratio 2. available Fe (mg kg-1) 3. As, Cd, Cr, Cu, Hg, Ni Pb (mg/kg-1)
Czech Republic	1	P	-	1. base saturation 2. exchangeable bases, exchangeable H
Denmark	No data	-	-	-
Estonia	4	FRO	KESE, SMI	1.soil texture class
Finland	No data	-	-	-
France	2	F	RMQS, BDGSF	1. major and trace elements 2. magnetic properties 3. organic matter fractions (MOP) 4. soil ADN bacteria and fungi 5. soil name and texture class 6. water regime 7. parent material
Germany	No data	-	-	-
Hungary	No data	-	-	-
Ireland	No data	-	-	-
Italy	No data	-	-	-
Latvia	4	P	-	1. Stoniness 2. base saturation (%), 3. H+ (mval 100g-1) 4. SOC stock 5. depth to water table
Lithuania	1	F	Mel_DR10LT	1. condition of soil drainage systems
Netherlands	No data	-	-	-
Norway	No data	-	-	-
Poland	No data	-	-	-
Portugal	No data	-	-	-
Slovakia	No data	-	-	-
Slovenia	No data	-	-	-
Spain	No data	-	-	-
Sweden	3	P	-	1. buffer capacity (tonnes CaO ha-1 to raise pH by one unit) 2. target pH (optimal pH for crop production
Switzerland	No data	-	-	-
Turkey	No data	-	-	-
United Kingdom	4	FRP	NSI_Profile, NSISC88, NSISC09,AFBI 5K	1. Soil Structure (the shape, size and degree of development of the aggregation) 2. Matrix colour (Munsell)
%*				40

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP3.1-14. Threats, data availability in participating countries

SP3. Threats 1-14	databases			SP3.1	SP3.2	SP3.3	SP3.4	SP3.5	SP3.6	SP3.7	SP3.8	SP3.9	SP3.10	SP3.11	SP3.12	SP3.13	SP3.14
Country	Relevant for topic	data policy	for EJP or freely	Soil erosion by water	Soil erosion by wind	Decline in SOM in peatsoils	Decline in SOM in mineral soils	Compaction, stucture degradation	Soil sealing	Pollution with potentially toxic elements	Pollution with organic substances	Acidification	Salinisation and alkalinisation	Deserti-fication	Flooding	Landslide	Decline in soil biodiversity
Austria	3	P	-	3	1	no	1	1	no	no	no	no	no	no	no	1	no
Belgium Flanders	1	F	PSEMF	1	no	no	no	1	1	1	1	no	no	no	1	1	no
Belgium Wallonia	No data	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Czech Republic	2	P	-	no	no	no	no	no	no	2	1	no	no	no	no	no	no
Denmark	2	P	-	no	no	no	1	no	no	no	no	1	no	no	no	no	no
Estonia	3	FR	SM,KESE, SMI	3	no	2	2	2	no	2	no	2	no	no	no	no	1
Finland	1	P	-	no	no	no	1	no	no	no	no	no	no	no	no	no	no
France	3	FP	RMQS, BDGSF	1	no	no	no	1	no	2	1	2	1	no	no	no	no
Germany	1	F	BZE_LW	no	no	no	no	1	no	no	no	no	no	no	no	no	no
Hungary	2	P	-	no	no	1	1	no	no	no	1	1	1	1	1	no	no
Ireland	3	FP	Tellus GSI	no	no	no	no	1	no	3	no	1	no	no	no	no	no
Italy	4	FP	SISI, PPD, NS	4	no	no	no	no	no	no	no	no	no	no	4	4	no
Latvia	5	PO	-	2	1	no	1	1	no	2	no	2	no	no	no	no	1
Lithuania	1	F	DR10LT	yes	1	no	no	1	1	no	no	1	no	no	no	no	no
Netherlands	No data	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Norway	4	F	NSS(S), AVRO, FHM	2	no	no	no	no	no	no	no	no	no	no	1	1	no
Poland	4	FRP	MChGO, MonFrm	no	no	no	1	no	no	1	1	4	no	no	no	no	no
Portugal	1	P	-	no	no	no	1	1	no	1	no	1	no	no	no	no	no
Slovakia	2	FP	BPEJ	2	1	1	1	2	no	1	no	1	1	no	no	no	no
Slovenia	No data	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Spain	4	P	-	no	1	no	no	1	no	2	1	no	no	no	no	no	1
Sweden	3	FP	SOILCOM ErRM	2	no	no	no	1	no	no	no	no	no	no	no	no	no
Switzerland	1	F	SWISOIL	no*	no*	no*	no*	no*	no*	no*	no*	no*	no*	no*	no*	no*	no*
Turkey	1	P	-	no	no	no	1	no	no	1	no	1	1	no	no	no	no
United Kingdom	9	FRP	NSISC88, NSISC09, AFBI 5K, UKSHPS, TELLUS A,B,XRF	2	2	no	no	no	no	9	2	no	no	no	no	no	no
Total number of databases	60			22	7	4	11	14	2	27	8	17	4	1	7	7	3
%*	92			40	24	12	40	48	8	44	28	44	16	4	16	16	12

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The remark no* (from Partner Switzerland) is: Threats are probably covered in publications, but not as datasets.

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

SP4. Soil classification and fertility, data availability in participating countries

SP4. 1-4. soil classification	databases			SP4.1	SP4.2	SP4.3	SP4.4		
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**soil type, national class**	**soil type, international class**	**soil fertility**	**data for status of soil biodiversity/ health**	**Name of national classification**	**Name of international classification**
Austria	5	FPO	eBOD	5	no	no	no	Austrian Soil Classification (Fink), Austrian Soil Clasification 2000/2011	-
Belgium Flanders	2	F	DOV, SMF	2	2	1	no	Belgian soil classification	WRB 2006 and WRB 2014
Belgium Wallonia	1	P	-	1	no	no	no	-	-
Czech Republic	3	FP	BPEJ	3	1	1	no	Taxonomic classification system of soils of the Czech Republic	WRB2006
Denmark	3	RP	DDSM	2	2	no	no	Danish Pedological Soil Classification System, Danish Soil Texture Classification	FAO, FAO1990, FAO1998, Soil Taxonomy, Soil Taxonomy 2nd Ed, WRB1994, WRB 1998
Estonia	4	FRO	SM,KESE, SMI	4	1	3	1	Estonian Soil Classification	WRB2006, 2014, 2015
Finland	2	P	-	2	no	2	2	-	-
France	10	FP	RMQS, BDGSF	2	6	3	1	Référentiel Pédologique 1995, 2008	WRB2006, FAO modified by CEC 1985
Germany	1	F	BZE_LW	1	1	no	no	AD Hoc AG Boden (2005)	WRB2006
Hungary	2	P	-	2	no	no	no	Hungarian Genetic Soil Classification	
Ireland	2	PO	-	2	2	no	no	Irish SIS	WRB2015
Italy	5	FP	SISI, PPD, NS	no	5	no	no	-	WRB2006
Latvia	10	FPRO	DSD, LLU	10	1	2	1	-	WRB2014
Lithuania	2	F	DR10LT, Dirv_vert	1	1	1	no	LTDK-99 according by WRB2006	WRB2006
Netherlands	7	FO	BRO: BHR-p, SFR, SGM	3	no	no	4	Dutch classification	-
Norway	3	F	NSS(S)	2	3	no	no	Based on WRB2014	WRB2014
Poland	3	FRP	MChGO, MonFrm	1	no	3	no	-	-
Portugal	3	FP	INFSOL, PROSOL	1	3	no	no	Portuguese classification Cardoso, J.C., 1974.	WRB2006
Slovakia	4	FP	BPEJ, PM400	4	3	1	no	Morphogenetic soil classification system of the Slovak Republic	WRB 2014(2015), FAO 1974
Slovenia	1	F	SPSLO	1	1	no	no	Slovenian soil classification	FAO
Spain	8	P	-	no	2	8	2	-	WRB2006; Soil Taxonomy (Soil Survey Staff, 1999), USDA
Sweden	3	FP	SOILCOM	4	3	no	no	National soil texture classification SS-ISO 11277	International soil texture classification according to USDA
Switzerland	1	F	SWISOIL	1	no	no	no	-	-
Turkey	1	P	-	1	no	1	no	Based on USDA Soil taxonomy	-
United Kingdom	7	FRP	NSISC88, NSISC09, NSMSC, NSMNI, AFBI 5K	7	2	no	no	Soil Classification of England and Wales (Avery, 1980), Soil Classification of Scotland (2013)	WRB2006, 2014
Total number of databases	93			62	39	26	11		
%*	96			92	68	44	24		

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

SP5. Soil functions and services, data availability in participating countries

SP5. 1-5 soil functions/services	databases			SP5.1	SP5.2	SP5.3	SP5.4	SP5.5	
Country	Relevant for topic	data policy	for EJP or freely	Water storage capacity	Plant productivity potential	Carbon sequestration potential	Biodiversity potential	other	mention in other
Austria	No data	-	-	-	-	-	-	-	
Belgium Flanders	No data	-	-	-	-	-	-	-	
Belgium Wallonia	No data	-	-	-	-	-	-	-	
Czech Republic	No data	-	-	-	-	-	-	-	
Denmark	No data	-	-	-	-	-	-	-	
Estonia	2	FO	Soil_map	1	no	no	no	1	Soil bonity
Finland	No data	-	-	-	-	-	-	-	
France	3	FP	RMQS	1	no	1	no	1	Trace elements contents
Germany	No data	-	-	-	-	-	-	-	
Hungary	2	P	-	1	-	-	1	-	
Ireland	1	F	Tellus GSI	no	no	1	no	no	
Italy	No data	-	-	-	-	-	-	-	
Latvia	No data	-	-	-	-	-	-	-	
Lithuania	1	F	DR10LT	1	1	1	1	no	
Netherlands	2	O	-	1	no	1	no	no	
Norway	No data	-	-	-	-	-	-	-	
Poland	1	R	MonFrm	no	1	no	no	no	
Portugal	2	F	INFSOL, PROSOL	2	no	no	no	no	
Slovakia	No data	-	-	-	-	-	-	-	
Slovenia	No data	-	-	-	-	-	-	-	
Spain	4	P	-	1	1	1	2	2	DOC, FDA, DHA
Sweden	1	P	.	no	no	no	1	no	
Switzerland	No data	-	-	-	-	-	-	-	
Turkey	1	P	-	no	no	1	no	no	
United Kingdom	No data	-	-	-	-	-	-	-	
Total number of databases	20			8	3	6	5	4	
%*	48			28	16	24	16	12	

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Annex 4 Soil management data

Soil management data availability in EJP SOIL countries (2020), where "1" indicates that there is data for that practice

Stocktake of data for soil management systems																											
THEMATIC GROUPs	Management practices (MG) – available **data or maps** are about *(concerning WP2 suggestions also)* :		Austria	Belgium FL	Belgium WALL	Czech Republic	Denmark	Estonia	Finland	France	Germany	Hungary	Ireland	Italy	Latvia	Lithuania	Netherlands	Norway	Poland	Portugal	Slovakia	Slovenia	Spain	Sweden	Switzerland	Turkey	United Kingdom	**SUM**
MG1. Soil tillage, farming system	MG1.1	Conventional/organic/other farming	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1	0	0	1	20
	MG1.2	Precision farming	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	4
	MG1.3	Tillage(conventional/ reduced/ strip-till/ no-tillage)	1	0	1	0	1	1	0	1	1	1	0	0	1	0	1	1	1	0	0	1	1	1	0	0	1	15
	MG1.4	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2
MG2. Crop and cropping system	MG2.1	Crop (data for main crop)	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	22
	MG2.2	Cropping system (e.g. monoculture, kind of rotation etc.)	1	0	1	0	1	1	0	1	1	0	0	0	1	1	1	0	1	0	1	0	1	1	0	0	0	13
	MG2.3	*other, please specify*	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3
MG3. Soil nutrient management and plant protection	MG3.1	Fertilization (mineral/organic/both)	1	1	1	0	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	0	0	1	18
	MG3.2	Microbiological preparations	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	3
	MG3.3	Pesticides	1	0	1	0	0	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	1	10
	MG3.4	*Soil conditioners related to soil /plant health , other than 3.2*	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
	MG3.5	*other, please specify*	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	4
MG4. Water management and related protection	MG4.1	Irrigation	1	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	1	0	1	1	0	0	0	1	11
	MG4.2	Drainage	1	0	1	1	1	1	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	10
	MG4.3	Soil conditioners related to water	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	MG4.4	*other, please specify*	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	3
MG5. Measures to control soil erosion	MG5.1	Terraces (wall, bench, ridges or raised beds, others)	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3
	MG5.2	Windbreak hedges	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	3
	MG5.3	Runoff water management systems (channels, etc)	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0		0	5
	MG5.4	Buffer strips	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	6
	MG5.5	In field erosion control measures (e.g. micro-dams between ridges,...)	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	MG5.6	Small scale buffering infrastructure (retention ponds, dams,...)	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2
	MG5.7	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
MG6. Permanent crops SSM	MG6.1	Green manuring	0	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0	1	0	0	0	1	1	0	0	0	9
	MG6.2	Cover crops	1	1	0	0	0	0	1	1	1	1	0	0	1	0	1	0	1	0	1	1	0	1	0	0	0	12
	MG6.3	Mulching	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	5
	MG6.4	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MG7. Other management practices	MG7.1	Amelioration (other than drainage M	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0	5
	MG7.2	Greenhouses	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	5
	MG7.3	*other, please specify*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
		SUM	12	8	12	3	6	8	3	15	14	10	0	7	20	10	7	5	8	3	8	5	12	13	0	0	10	199

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG 1. Soil tillage, farming system, data availability in participating countries

	MG 1. Soil tillage, farming system						
	databases			MG 1.1	MG1.2	MG 1.3	MG 1.4
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Conventional/organic/ other farming**	**Precision farming**	**Tillage(conventional/ reduced/ strip-till/ no-tillage)**	**other**
Austria	4	PO	-	3	1	4	0
Belgium Flanders	1	O	-	1	0	0	0
Belgium Vallonia	2	P	-	2	0	1	0
Czech Republic	1	O	-	1	0	0	0
Denmark	2	P	-	1	0	1	0
Estonia	3	FR	KESE, SMI	3	0	2	0
Finland	2	P	-	2	0	0	0
France	1	F	RMQS	1	0	1	0
Germany	2	F	BZE_LW, DESTATIS	2	0	2	0
Hungary	2	P	-	2	1	1	0
Ireland	No data	-	-	-	-	-	-
Italy	1	R	RICA	1	0	0	0
Latvia	7	P	-	6	2	5	1
Lithuania	1	F	Mel_DR10LT	0	0	0	1
Netherlands	3	O	-	2	0	3	0
Norway	2	FR	AR5, DirPay	1	0	2	0
Poland	1	R	MonFrm	1	0	1	0
Portugal	No data	-	-	-	-	-	-
Slovakia	1	P	-	1	0	0	0
Slovenia	1	F	LPIS	1	0	1	0
Spain	10	P	-	10	0	10	0
Sweden	4	FP	SOILCOM	4	1	3	0
Switzerland	No data	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-
United Kingdom	1	F	NI_AgriCEN	1	0	1	0
Total number of databases	52			46	5	38	2
%*	72			80	16	60	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG 2. Crop and cropping system, data availability in participating countries

	MG 2. Crop and cropping system					
	databases			MG 2.1	MG 2.2	MG 2.3
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Crop (data for main crop)**	**Cropping system (e.g. monoculture, kind of rotation etc.)**	**other**
Austria	4	PO	-	4	4	0
Belgium Flanders	2	F	Landbou., Groeicurve	1	0	2
Belgium Vallonia	2	P	-	1	1	1
Czech Republic	1	P	-	1	0	0
Denmark	2	P	-	2	2	0
Estonia	3	FR	KESE, SMI	2	2	0
Finland	2	P	-	2	0	0
France	1	F	RMQS	1	1	0
Germany	2	F	BZE_LW, DESTATIS	2	2	0
Hungary	3	P	-	3	0	0
Ireland	No data	-	-	-	-	-
Italy	1	R	RICA	1	0	0
Latvia	7	P	-	5	6	0
Lithuania	1	F	DR10LT	1	1	0
Netherlands	3	O	-	3	3	0
Norway	1	F	DirPay	1	0	0
Poland	1	R	MonFrm	1	1	0
Portugal	2	F	COS, PUBIRRAR	2	0	0
Slovakia	2	P	-	2	1	0
Slovenia	1	F	LPIS	1	0	0
Spain	8	P	-	3	8	0
Sweden	4	FP	SOILCOM	4	3	0
Switzerland	No data	-	-	-	-	-
Turkey	No data	-	-	-	-	-
United Kingdom	4	FP	CROME, CEHLCM, NI_AgriCEN	3	0	1
Total number of databases	57			46	35	4
%*	80			88	52	12

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents*

MG3. Soil nutrient management and plant protection, data availability in participating countries

	MG3. Soil nutrient management and plant protection							
	databases			MG3.1	MG3.2	MG3.3	*MG3.4*	*MG3.5*
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Fertilization (mineral/organic/both)**	**Microbiological preparations**	**Pesticides**	**Soil conditioners related to soil /plant health other than 3.2**	**other**
Austria	4	PO	-	3	0	1	0	1
Belgium Flanders	1	F	Bemest.	1	0	0	0	0
Belgium Vallonia	3	P	-	3	0	2	0	0
Czech Republic	No data	-	-	-	-	-	-	-
Denmark	1	P	-	1	0	0	0	0
Estonia	2	FR	KESE, SMI	2	0	1	1	0
Finland	No data	-	-	-	-	-	-	-
France	1	F	RMQS	1	1	1	0	0
Germany	2	F	BZE_LW, DESTATIS	2	0	0	0	0
Hungary	2	P	-	2	0	1	0	0
Ireland	No data	-	-	-	-	-	-	-
Italy	1	R	RICA	1	0	1	0	0
Latvia	7	P	-	7	2	4	0	1
Lithuania	No data	-	-	-	-	-	-	-
Netherlands	4	O	-	4	0	0	0	0
Norway	1	F	DirPay	1	0	0	0	0
Poland	1	R	MonFrm	1	0	0	0	0
Portugal	1	P	-	1	0	0	0	0
Slovakia	1	F	BPEJ	1	0	0	0	0
Slovenia	No data	-	-	-	-	-	-	-
Spain	10	P	-	10	1	2	4	2
Sweden	2	P	-	2	0	1	0	0
Switzerland	No data	-	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-	-
United Kingdom	6	FP	SCOPES, Comp_Risk, NI_AgriCEN, AFBI_Pest	2	0	3	0	1
Total number of databases	50			45	4	17	5	5
%*	68			72	12	40	8	12

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG 4. Water management and related protection, data availability in participating countries

	MG 4. Water management and related protection						
	databases			MG 4.1	MG 4.2	MG 4.3	MG 4.4
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Irrigation**	**Drainage**	**Soil conditioners related to water protection**	**other**
Austria	4	PO	-	3	2	0	1
Belgium Flanders	No data	-	-	0	0	0	0
Belgium Wallonia	1	P	-	0	1	0	0
Czech Republic	1	P	-	0	1	0	0
Denmark	2	RP	DDSM	0	2	0	0
Estonia	3	FR	KESE, SMI	0	3	0	0
Finland	No data	-	-	-	-	-	-
France	2	F	RMQS, BDGSF	2	2	0	0
Germany	2	F	BZE_LW, DESTATIS	2	2	0	0
Hungary	1	P	-	1	0	0	0
Ireland	No data	-	-	-	-	-	-
Italy	1	R	RICA	1	0	0	0
Latvia	5	P	-	2	4	0	0
Lithuania	1	F	Mel_DR10LT	1	1	0	1
Netherlands	No data	-	-	-	-	-	-
Norway	No data	-	-	-	-	-	-
Poland	No data	-	-	-	-	-	-
Portugal	1	F	PUBIRRAR	1	0	0	0
Slovakia	No data	-	-	-	-	-	-
Slovenia	1	F	MSYS	1	0	0	0
Spain	2	P	-	2	0	0	0
Sweden	2	P	-	0	2	0	0
Switzerland	No data	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-
United Kingdom	8	F	Eng_WS, NZV (Eng, Wales, Scot, NI), Eng_Fl_Risk, AWC_Scot, Runoff_Scot	1	0	0	7
Total number of databases	37			17	20	0	9
%*	60			44	40	0	12

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG5. Measures to control soil erosion, data availability in participating countries

	MG5. Measures to control soil erosion									
	databases			MG 5.1	MG 5.2	MG 5.3	MG 5.4	MG 5.5	MG 5.6	MG 5.7
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Terraces (wall, bench, ridges or raised beds, others)**	**Windbreak hedges**	**Runoff water management systems (channels, etc)**	**Buffer strips**	**In field erosion control measures (e.g. micro-dams between ridges,...)**	**Small scale buffering infrastructure (retention ponds, dams,...)**	**other**
Austria	No data	-	-	-	-	-	-	-	-	-
Belgium Flanders	1	F	Uitgevoerde.	0	0	0	1	1	1	0
Belgium Wallonia	2	P	-	0	0	2	2	2	0	0
Czech Republic	No data	-	-	-	-	-	-	-	-	-
Denmark	No data	-	-	-	-	-	-	-	-	-
Estonia	No data	-	-	-	-	-	-	-	-	-
Finland	No data	-	-	-	-	-	-	-	-	-
France	1	F	RMQS	1	1	1	1	0	0	0
Germany	1	F	BZE_LW	1	0	1	0	0	0	0
Hungary	No data	-	-	-	-	-	-	-	-	-
Ireland	No data	-	-	-	-	-	-	-	-	-
Italy	No data	-	-	-	-	-	-	-	-	-
Latvia	2	P	-	0	1	0	1	0	0	0
Lithuania	2	F	DR10LT, Mel_DR10LT	0	0	2	0	0	1	0
Netherlands	No data	-	-	-	-	-	-	-	-	-
Norway	No data	-	-	-	-	-	-	-	-	-
Poland	No data	-	-	-	-	-	-	-	-	-
Portugal	No data	-	-	-	-	-	-	-	-	-
Slovakia	1	P	-	1	1	0	1	0	0	0
Slovenia	No data	-	-	-	-	-	-	-	-	-
Spain	No data	-	-	-	-	-	-	-	-	-
Sweden	3	P	-	0	0	1	3	0	0	0
Switzerland	No data	-	-	-	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-	-	-	-
United Kingdom	2	FRP	GB_BS_Er., ErRisk_Scot	0	0	0	0	0	0	2
Total number of databases	15			3	3	7	9	3	2	2
%*	36			12	12	20	24	8	8	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

** Percentage expression of positive results in questionnaires obtained from 25 respondents*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG 6. Permanent crops for sustainable soil management (SSM), data availability in participating countries

	MG 6. Permanent crops SSM						
	databases			MG 6.1	MG 6.2	MG 6.3	MG 6.4
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Green manuring**	**Cover crops**	**Mulching**	**other**
Austria	1	P	-	0	1	0	0
Belgium Flanders	1	F	Landbou.	0	1	0	0
Belgium Vallonia	2	P	-	2	0	0	0
Czech Republic	No data	-	-	-	-	-	-
Denmark	No data	-	-	-	-	-	-
Estonia	No data	-	-	-	-	-	-
Finland	2	P	-	0	2	0	-
France	1	F	RMQS	1	1	0	0
Germany	2	F	BZE_LW, DESTATIS	2	2	1	0
Hungary	1	P	-	1	1	1	-
Ireland	No data	-	-	-	-	-	-
Italy	No data	-	-	-	-	-	-
Latvia	4	P	-	2	2	2	1
Lithuania	No data	-	-	-	-	-	-
Netherlands	2	O	-	2	1	0	0
Norway	No data	-	-	-	-	-	-
Poland	1	R	MonFrm	1	1	1	0
Portugal	No data	-	-	-	-	-	-
Slovakia	1	P	-	0	1	0	0
Slovenia	1	F	LPIS	0	1	0	0
Spain	2	P	-	1	0	1	0
Sweden	3	P	-	2	3	0	0
Switzerland	No data	-	-	-	-	-	-
Turkey	No data	-	-	-	-	-	-
United Kingdom	No data	-	-	-	-	-	-
Total number of databases	24			14	17	6	1
%*	48			36	48	20	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

* Percentage expression of positive results in questionnaires obtained from 25 respondents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

MG 7. Other management practices, data availability in participating countries

	MG 7. Other management practices					
	databases			MG 7.1	MG 7.2	MG 7.3
Country	**Relevant for topic**	**data policy**	**for EJP or freely**	**Amelioration (other than drainage MG4.2)**	**Greenhouses**	**other**
Austria	No data	-	-	-	-	-
Belgium Flanders	No data	-	-	-	-	-
Belgium Wallonia	No data	-	-	-	-	-
Czech Republic	No data	-	-	-	-	-
Denmark	No data	-	-	-	-	-
Estonia	No data	-	-	-	-	-
Finland	No data	-	-	-	-	-
France	No data	-	-	-	-	-
Germany	2	F	BZE_LW, DESTATIS	1	1	0
Hungary	No data	-	-	-	-	-
Ireland	No data	-	-	-	-	-
Italy	1	R	RICA	1	1	0
Latvia	2	FP	DigProf	0	1	1
Lithuania	2	F	DR10LT, Mel_DR10LT	2	1	0
Netherlands	No data	-	-	-	-	-
Norway	1	F	DirPay	0	1	0
Poland	No data	-	-	-	-	-
Portugal	No data	-	-	-	-	-
Slovakia	No data	-	-	-	-	-
Slovenia	No data	-	-	-	-	-
Spain	No data	-	-	-	-	-
Sweden	2	FP	SOILCOM	2	0	0
Switzerland	No data	-	-	-	-	-
Turkey	No data	-	-	-	-	-
United Kingdom	No data	-	-	-	-	-
Total number of databases	10			6	5	1
%*	24			16	20	4

Data policy : F-freely ; R- freely for research purpose EJP SOIL; P-permission; O- other

The numbers in the table show the sum of databases that relevant to the given topic, indicated by the 25 respondents.

* Percentage expression of positive results in questionnaires obtained from 25 respondents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Annex 5 Global Soil Partnership CountrySIS survey

The Global Soil Partnership of FAO has launched a questionnaire to collect information on soil data available at country level, named CountrySIS survey (https://forms.gle/X6N2G4WX86VYk8tn9). This is how it is officially described CountrySIS survey by GSP: "The survey aims to assess soil databases and information systems currently existing on the national level, in order to plan global activities according to the capacities and needs of the countries. The PDF preview of the survey questions can be accessed here: https://drive.google.com/open?id=1rboHEXP-7LO3mY9fFhivk3Gc80lcqBGa. The survey starts from questions, related to the soil property database. If the database is a part of a soil information system (SIS), it continues to describe the SIS (if not, the SIS section can be skipped). The last section is dedicated to the soil monitoring system (if it is present). A soil property database is a collection of measured values of soil properties organized in a digital format so that it can be easily accessed, managed and updated. The data may be associated with soil profiles (soil profile database) or with the mapping units. A Soil Information System (SIS) is a geographic information system (GIS) for the capture, storage, management, processing and display of soil-related data from original sources. A soil property database may serve as the main component of the SIS. SIS can also include non-soil data (such as climate or land use) to support land-management decision making. A Soil Monitoring System is based on the regular repetitive soil sampling aimed at observing the change of soil properties over time in order to control soil quality and address soil degradation."

In the first months of EJP-SOIL WP6 activity, after asking permission to GSP, we have distributed the questionnaire to EJP-SOIL partners. Some of them had already answered to the questionnaire on 2018.

Who answered to CountrySIS, among EJP-SOIL countries?

16 countries of EJP-SOIL consortium have answered, with answers given by the EJP-SOIL institutions (green coloured in the table). For Estonia, Portugal, and Slovenia we have answers of 2018, but of institutions not part of EJP-SOIL (yellow coloured in the table). For Switzerland we have received an answer from an institution external to EJP SOIL consortium but that collaborates with the EJP SOIL swiss partner (green coloured). For Belgium, Germany, Hungary, Italy, Spain, and United Kingdom we have either answer from EJP-SOIL institutions (green coloured) and from other institutions (sky blue coloured). In the report we are publishing only the answers received from EJP-SOIL institutions.

YEAR	Country	Institution / affiliation	EJP-SOIL partner
2020	Austria	BIOS-BFW	Y
2020	Austria	BIOS- EAA	Y
2018	Belgium	EV-ILVO	Y
2018	Belgium	OVAM (Public Waste Agency Flanders)	N
2020	Denmark	AU	Y
2018	Estonia	Estonian Environment Agency	N
2018	Finland	LUKE	Y
2020	France	INRAE	Y

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

2018	Germany1	THÜNEN	Y
2018	Germany2	THÜNEN	Y
2018	Germany	Federal Environment Agency (UBA)	N
2020	Hungary1 to 3	MTA-ATK	Y
2020	Hungary4	MTA-ATK	Y
2018	Hungary	National Food chain Safety office	N
2018	Italy	Università di Pavia	N
2018	Italy	CREA	Y
2020	Latvia1	UL	Y
2020	Latvia2&3	UL	Y
2020	Lithuania	LAMMC	Y
2021	Netherlands	WR	Y
2020	Norway	NIBIO	Y
2020	Poland	IUNG	Y
2018	Portugal	Instituto Superior Técnico	N
2018	Portugal	University of Evora	N
2018	Slovakia	NPPC	Y
2018	Slovenia	Ministry of the Environment and Spatial Planning, Slovenian Environment Agency	N
2018	Spain	Universitat de Barcelona	N
2018	Spain	CSIC	Y
2018	Spain	ule	N
2018	Spain	Tracasa/Navarra Government. Soil data is only from Navarra, not whole Spain	N
2018	Spain	Institut Cartogràfic i Geològic de Catalunya	N
2021	Switzerland	SFOE-AGS (Swiss Federal Office for the Environment, contacted by AGS)	N (Y)
2020	Turkey	TAGEM	Y
2020	United Kingdom (Northern Ireland)	AFBI	Y
2018	United Kingdom	Natural England	N
2018	United Kingdom	Centre for Ecology and Hydrology	N
2018	United Kingdom	James Hutton Institute	N
2018	United Kingdom	Cranfield University	N

Questions related to the Digital Database of Soil Properties

The following questions were made:

Does your country have digital databases of soil properties?
Name and affiliation of the database
Is your institution / database specialized on certain data (e.g. soil maps of a certain scale, soil profiles, hazard maps, other)? Please specify further.
Who are the users of the database?
For what purposes the data is used in your country?
Is the data publicly accessible?
Is the data distributed with open license?
What is the main spatial unit of the database?
How many soil profiles are there in the database?
How many topsoil samples are there in the database?
Please describe the common site (plot) description and method of sample extraction
Are all profiles / sampling locations georeferenced?
Can you identify precisely the coordinate reference system (CRS) used in the database? (e.g. WGS 84 / UTM Zone 36N, EPSG:32636)
Please indicate the number of samples in the database coming from before 1960
Please indicate the number of samples in the database coming from 1960-1990
Please indicate the number of samples in the database coming from 1990-2010
Please indicate the number of samples in the database coming from after 2010
Has there been any recent update of the database (or plan for its update)?
Does your institution use a common list of soil properties?
Part 1: General Properties
How many profiles have information on all the parameters selected above?
Part 2: Plant Nutrients
How many profiles have accompanying data on plant nutrients?
Part 3: Soil Salinity

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement N° 862695

How many profiles have accompanying data on soil salinity?
Part 4: Soil Pollution / Contamination
How many profiles have accompanying data on soil pollution / contamination?
What is the format of the database?
What kind of metadata is included in the database?
What metadata standards are used?
Please provide a link to documentation (papers/reports) describing the database structure (optional)
Please describe the database structure (optional)
Do you have a quality control for the data in the database?
Please describe the quality control/assessment procedure
Please indicate main challenges and capacity lacks in developing and maintaining a digital databases of soil properties

The answers received are reported in the tables 1 to 4.

Table 1 (green color for data publicly available, red for not publicly available, light blue for intermediate accessibility, blue-gray for Turkey which has a digital soil database in process of establishing, grey for Spain which does not have a digital database of soil properties)

Country	Institution / affiliation	Does your country have digital databases of soil properties?	Name and affiliation of the database	Is your institution / database specialized on certain data?	Who are the users of the database?	For what purposes the data is used in your country?	Is the data publicly accessible?	Is the data distributed with open license?
Austria1	BIOS-BFW	Yes	Agricultural Soil Map of Austria, BFW	Soil Map of agricultural soils 1 : 25.000, area related data, derived thematic maps (soil type, soil type group, parent material, soil depth, water balance,	farmers, schools, universities, public	efficient cultivation (e.g. fertilization,	Yes	for the most part

				permeability, humus form, texture, humus content, content of carbonates, chemical reaction, value	authorities, agricultural associations, freelance civil engineers, private persons	irrigation), natural hazard management (e.g. erosion, flood), pre soil sampling information (e.g. application of sewage sludge), cultural and constructional technics (e.g environmental impact assessements), land		
Austria2	BIOS-EAA	Yes	BORIS Soil Information System	Point related data on site- and soil profile description (Land use, Soil type...) as well as analytical data: Soil Physics, nutrients, carbon, contaminants: heavy metals, organic pollutants... Over 1.2 Mio data for 600 soil related parameters for over 10000	Universities, scientists, experts in national and federal institutions, experts & specially certified engineers	Strategic Environmental Assessment (SEA) & Environmental Impact Assessment (EIA), national and	Site and profile information as well as maps are open to the public, analytical values are provided upon approval of the particular data owners	No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

					working in the field of Strategic Environmental Assessment (SEA) & Environmental Impact Assessment (EIA)	European soil related projects, reporting for European and national strategies for soil protection, soil assessment for different purposes		
Belgium	EV-ILVO	Yes	Bodemdatabank DOV	soil data except data about contamination and nutrients N and P	anybody	diverse purposes	a selection of the data	a selection of the data
Denmark	AU	Yes	Danish Soil Profile Database	The database contains information on soil profiles. Our institution focuses on maps of soil properties at national scale.	Researchers at Aarhus University and Copenhagen University	For mapping, for calibration of spectroscopic models, to study soil processes and properties	No	No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Finland	LUKE	Yes	Biosoil data and Valse data	Soil profiles and chemical analysis of those	Researchers from LUKE	Research	No	No
France	INARE	Yes	Donesol INRAE-InfoSol	INRAE InfoSol has a long tradition in soil survey, monitoring and mapping (including DSM). It acts as the National coordinator of survey and monitoring programmes for France under the supervision of GIS Sol (www.gissol.fr). It is the national center for soil data.	Universities, students, researchers, institutes, Ministries, private companies, agricultural and forestry development organizations	Databases are used to provide information about soil spatial distribution, properties and quality. It is also used as input parameters for spatial modelling and analyses (for agricultural and environmental issues).	It depends on the ownership of the data but part is publicly accessible (point data with exact coordinates are not)	For publicly available data, the licence is the French open data licence compatible with creative commons licence (data available through INRAE data portals data.inrae.fr and agroenvgeo.inra.fr)
Germany 1	THUNEN	Yes	BZE Wald	**forest** soil maps (Kyoto protocol), 1:1.000.000, nutrition deficiency maps, soil acidification	scientist, foresters, policy maker	Condition of Forests and especially forests soil in Germany	via -Email	in 2019 all aggregated data will be open, 2018 only 1. inventory

Germany 2	THUNEN	Yes	Agricultural Soil Inventory	focus on **soil organic carbon**	internal use only	establish a **soil monitoring** on stocks of soil organic carbon development	not yet available due to running project	No
Hungary 1	MTA-ATK	Yes	DIGITAL KREYBIG SOIL INFORMATION SYSTEM (DKSIS), Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research	The Institute has a long tradition in soil survey and mapping, mainly as a coordinator, think tank, initiator, legacy data manager and most recently as national Digital Soil Mapping center.	Universities, lecturers, students, researchers, institutes	The database is used as soil information parameters for spatial modelling and analyses	No	No
Hungary 2	MTA-ATK	Yes	Hungarian Soil Information and Monitoring System (SIMS), Directorate of Plant Protection and Soil	Hungarian Soil Information and Monitoring System (SIMS) was established for monitoring purposes. Certain characteristics were measured annually, every three or six years, however, slowly or non changeable soil properties were only measured once, in the fi	Universities, lecturers, students, researchers, institutes	The database is used as soil information parameters for spatial modelling and analyses	available by request	No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

		Yes	Conservation, National Food Chain Safety Office (NÉBIH)					
Hungary 3	MTA-ATK	Yes	TDR (Soil Degradation Subsystem of the Hungarian Environmental Information System), Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research	The Institute has a long tradition in soil survey and mapping, mainly as a coordinator, think tank, initiator, legacy data manager and most recently as national Digital Soil Mapping center.	Universities, lecturers, students, researchers, institutes.	The database is used as soil information parameters for spatial modelling and analyses.	No	No
Hungary 4	MTA-ATK	Yes	Hungarian Detailed Soil Hydrophysical Database (acronym of the Hungarian name of the dataset: MARTHA), Institute for	Our institute has a long tradition in soil survey and mapping, mainly as a coordinator, think tank, initiator, legacy data manager and most recently as a national Digital Soil Mapping Center.	Universities, research institutes, lecturers, researchers, MSc and PhD students.	The data is used for mapping soil properties and functions, deriving relationship between soil	No	No

			Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research			properties, environmental modelling.		
Italy	CREA	Yes	Soil Information System of Italy, CREA-AA	**All** of them	Researchers, anybody else under request. Data used for soil mapping at national scale.	Research	The data owned by CREA can be shared with anybody. Authorization to other soil data owners is needed for data stored from other data owners.	Yes
Latvia1	UL	Yes	Crop management information system of Latvia	No	Users are employees of the State Plant Protection Service.	Data are used for agricultural purposes, agro-chemical soil research, as well as collected and thus used for agricultural and environmental policy	No	No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

						planning purposes.		
Latvia2	UL	Yes	Forest soil monitoring program, LSFRI Silava	Yes, we are elaborating soil maps for National GHG inventory. We are using other data sources for elaboration of the database.	National GHG inventory team	National GHG inventory in LULUCF sector	Data are elaborated particularly for GHG inventory to be used with Yasso model and emission accounting in organic soils.	No
Lithuania	LAMMC	Yes	Dirv_DR10LT (State enterprise State Land Fund)	State Land Fund is specialized on soil maps of a certain scale, soil profiles, spatial datasets (reclamation status and sodden soils, limited land use areas, abandoned agricultural land).	Land managers, landowners and users, students and researchers, etc.	Land management and spatial planning.	Yes	Yes
Netherlands	WR	Yes	3 databases: National Key Registry of the Subsurface, BIS, Geochemical Atlas of the Netherlands	1:50.000 national soil class map, soil profile classification and description, detailed soil surveys, sampling	Consultancy companies and governmental organisations	Land use planning, agriculture, (ground and surface) water modelling and management, infrastructure, environmental	Yes	Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

						background values, other		
Norway	NIBIO	Yes	The Norwegian Soil Profile Database	The database is made for general purposes, and is used for producing different maps and statistics	Researchers, public management at national, regional and local level, and farmers	For agronomy and physical planning, risk assessments	Yes	Yes
Poland	IUNG	Yes	Soil - agricultural map	my institution is specialized in soil-agricultural maps at various scales, soil profile databases, acidity and SOM maps, trace metal maps	national and regional administration, science, agencies, spatial planning offices	delineation of regions for support, analysis of trends, implementation of policies, spatial planning	soil maps for certain regions in various scales are public available mainly to view online	No
Slovakia	NPPC	Yes	Digital database of agricultural soils of Slovakia, affiliated with National Agricultural Land Food	The institute is national authority for agricultural soil information / the database consist of soil maps in variety of scales and data on soil profiles	government, municipalities, soil users, researchers, public	soil conservation according to the national environmental legislation, land use planning	Yes	only selected datasets

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

			Centre, Soil Science and Conservation Research Institute			including land reclamations, sustainable soil and land management		
Spain	CSIC	No						
Switzerland	SFOE-AGS	Yes	Nationale Bodendatenbank NABODAT ; operated by the Swiss confederation	All types of soil data are in our focus. See data model documentation published on https://www.nabodat.ch/index.php/de/fachapplikation/datenmodell (in German and French)	The principal users are experts of the cantonal / regional authorities for soil protection, as well as experts on the national level.	The main purpose of the database is to store and offer Switzerland's soil data in a digital and harmonized form. The data is used on the one hand as a knowledge base for the pedological properties of our (very	A subset of the data is available on https://www.nabodat.ch/index.php/de/service/bodendatensatz	The data that is publicly available follows the Aarhus convention rules. Therefore, it is distributed with an open licence

						heterogeneous) soils, and on the other hand as a tool for the execution of the soil protection laws, including the management of soil threats		
Turkey	TAGEM	In process of establishing	National Soil Information System (UTBS)	Yes	Public institutions, universities, researchers, policy makers, farmers, NGOs etc.	agricultural production, research, education, supporting related policies	under construction	No
United Kingdom (Northern Ireland)	AFBI	Yes	AFBI Soil Attributes Dataset	Not specialized as such, hold soil maps, profile and attribute data for Northern Ireland	Government - Agriculture, Environment, Planning; SMEs and	Agri-environmental analysis mainly. Planning with regard to	Maps are available at cost, digital data at 1:250,000 is publicly available. For commercial use data is available at cost under license.	Yes, depending on scale and if the data is for commercial use.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

					farmers to a lesser degree	environmental impact assessments.		

Table 2 (green color for data publicly available, red for not publicly available, light blue for intermediate accessibility, blue-gray for Turkey which has a digital soil database in process of establishing, grey for Spain which does not have a digital database of soil properties)

Country	Survey year	spatial unit of the database?	How many soil profiles	How many topsoil samples	Site (plot) description and method of sample extraction	Are georeferenced?	Coordinate reference system (CRS	Number of samples in the database				
								before 1960	from 1960-1990	from 1990-2010	after 2010	Recent update
Austria1	2020	both (mapping units and reference profiles)	12	12	defining mapping units using a hammer drill each 100 metres of catena, respectively occasion related, on a field map 1 : 10.000; digging profile pits to a depth of 1meter (if possible) for each mapping unit, horizon related description and sampling for an	the locations of the hammer drill surveys are only marked at the analog field maps, the locations of the reference profiles were conveyed to the analog maps, which were the basis for digitalization (accuracy about 20 meters), only very	WGS 84 / UTM Zone 32/33, EPSG:4326	300	45	2500	200	Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

						few of them where lo						
Austria2	2020	Point-based (soil profiles)	over 10000	over 10000	Site description is different in each study depending on the selected Parameters, the big soil Surveys follow - with some variance - national regulations. Sample extraction is done following national regulations - due to land use or purpose of the study -	Yes	EPSG: 31287 / MGI /Austria Lambert	none	about 600 - 700 sites / usually 2 - 3 different soil depths for analysis	8000 sites / usually 2 - 3 different soil depths for analysis	2000	Yes
Belgium	2018	point-based and polygon-based	7.762	16.301	diverse (pit and bore)	Yes	EPSG/31370	16.301 samples during period 1945 - 1980	16.301 samples during period 1945 - 1980	0 (but in 2018 samples will be incorporate d in the database)	0 (but in 2018 many samples will be incorporat ed in the database)	Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Denmark	2020	Point-based (soil profiles)	2476	2233, however we have additional topsoil samples (50,000) stored in other formats	Profile database: Bulk samples and ring samples extracted from pedogenetic horizons. Exact depth of the sample also noted. Additional data: Topsoil composites from 25 - 30 samples (40,000) or soil cores from five 20-cm depth intervals (10,000).	No	ED50 or ETRS1987, UTM Zone 32N	0	10,813 profile samples. 40,000 additional topsoil samples	1817 profile samples, 10,000 soil cores	86 profile samples	No
Finland	2018	Point-based (soil profiles)	521 for forests and 600 for agrisoils	around 1000	Check here: http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/15905/1/lbna24729enc.pdf	Yes	EPSG: 2393	0	around 1000 in total for forests and agrilands	around 1000 in total for forests and agrilands	For Valse there are around 3000	For agrilands yes
France	2020	Both point- and polygon-based, also raster data for predicted soil properties	73295 soil profiles, 98751 auger holes	around 50% of the soil profiles have soil analysis on topsoil	for recent soil profiles, the description follows the French norm for soil description (STIPA). Samples are taken in each horizon for analyses (mottles), cylinders are generally used for bulk density measurement when it is done	Not all	when it is georeferenced it is generally in WGS 84 for non overseas profiles	0.04%	0.56%	0.31%	0.08%	Yes
Germany1	2018	Point-based (soil profiles)	1800 for each inventory (8x8km grid)	1800 (for organic layer, 0-5, 5-10,10-30)	8 satellites in 10m distance from the centre, 7 depth	Yes	WGS84	0	1800	1800	0	Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Germany2	2018	Point-based (soil profiles)	3100	3100; 0-10 and 10-30 cm each	scientific site description in the field after German regulation (KA5); samples are taken as disturbed and undisturbed samples following a protocol which respects stoniness of the soil	Yes	UTM Zone 38N, UTM Zone 39N	0	0	0	3100	project is still running
Hungary1	2020	Both point- and polygon-based	Detailed profile descriptions: ~22,000 sampling sites (cca. 1 obs./4.23 km2), spatially transferred to an additional 250,000 locations (cca. 1 obs./0.37 km2).	Profile descriptions contain topsoil information as well.	Site description: position, depth of organic layer, land use (crop). Sampling: upper and lower depth, colour (not standardized, subjective names), structure.	Yes	HD72 / EOV, EPSG:23700	Detailed profile descriptions: ~22,000 sampling sites (cca. 1 obs./4.23 km2), spatially transferred to an additional 250,000 locations (cca. 1 obs./0.37 km2).	0	0	0	No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Hungary2	2020	Point-based (soil profiles)	1236	1236	Site description: position, aspect, erosion, vegetation. Sampling: upper and lower depth, depth of organic layer, depth of groundwater, parent material.	Yes	HD72 / EOV, EPSG:23700	0	0	1236 profiles, ~4500 samples	0	No
Hungary3	2020	X' pattern mixed sampling of an 5 ha representative quadrat of the parcel	1958	1958		Yes	WGS 84, EPSG:4326	0	0	0	6939	No
Hungary4	2020	Point-based (soil profiles)	3970	3495	Site description: position. Method of sample extraction: sampling depth followed the depth of the genetic horizons.	Yes	HD72/EOV, EPSG:23700	0	6083 soil samples of 1552 soil profiles	9059 soil samples of 2418 soil profiles	0	Yes
Italy	2018	Polygon-based (mapping units)	1420 public, 22860 all	67490	profiles and auger holes, soil derived profiles	Yes	WGS 84 LAT LONG	29	3874	38796	1743	Yes
Latvia1	2020	Polygon-based (mapping units)	None	159’984		Yes	Latvian LKS-92	None	None	Form 2006 - 2019. Samples: 159’984	Form 2006 - 2019. Samples: 159’984	No

Latvia2	2020	National forest inventory plot level information, multiple sources are used in soil characteristics	95	380	ICP Forests level II forest soil monitoring program	Yes	WGS84	No data available	No data available	950	1050	Yes
Lithuania	2020	Point-based (soil profiles)	107 618	198 896	Site (plot) description: landform and topography, parent material, land use and vegetation. After horizon delineation samples were taken from each horizon of soil profile up to 1.5-2.0 m depth.	Yes	Projected Coordinate System LKS1994Lithuania TM	The number of profiles - 30 669	The number of profiles - 65 530	The number of profiles - 11 302	The number of profiles - 117	Yes
Netherlands	2020	Both	332000	A couple of thousands	Dutch soil description using a Dutch soil auger	Yes	RD (Rijksdriehoek stelsel)	Little	60%	30%	10%	Yes
Norway	2020	Two different databases, one point-based and one polygon-based	Approx. 3 000	Approx. 3 000	Strategic sampling for the soil survey	Yes	Yes	None	Approx. 500 (not checked)	Approx. 2000 (not checked)	Approx. 500 (not checked)	Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Poland	2020	Polygon-based (mapping units)	10000 in a map database	225000	site description available only for limited number of samples, partly still not digitalized	Yes	WGS 84 or EPSG 2180	no such old samples	10000	50000	165000	Yes, but only regarding soil pH and SOM
Slovakia	2020	both soil profiles and mapping units	18059	18059	Description of position, topography, land cover and land use at the plot, excavated soil profiles, genetic and/or sequentional soil sampling up to 120cm, sampling into cylinders for soil phycical analyses.	Yes	S-JTSK (EPSG:5514), WGS84/UTM Zone 34 N (EPSG:32634)	0	66494	954	636	for some datasets yes
Switzerland	2020	Point-based (soil profiles)	Approx. 22’000 soil profiles. There is soil informatio n of a total of 59’000 sites. There are approx. 850’000 measurem ents stored in the database	There are approx. 39'000 topsoil samples (0-10, 0-20, 0-30 cm, and profil horizons 0-30 cm) of a total of approx. 78'000 samples	Most fixed-depth sampling is done 0-20 cm; except soil biological sampling (0..10 cm). The method of sample extraction varies according to type of sampling; e.g. sampling for soil pollution monitoring is different than sampling for mapping purposes. An old, but still valid documentation of soil pollution sampling is available here: https://www.bafu.admin.ch/bafu/de/home/themen/boden/publikationen-studien/publikationen/handbuch-	Yes	The standard Swiss coordinate reference system is used in NABODAT, LV95. Further info: https://www.swisstopo.admin.ch/de/wissen-fakten/geodaesie-vermessung/koordinaten/schweizer-koordinaten.html	Very few	Approx. 35'000 - precise number needs to validated	Approx. 30'000 - precise number needs to validated	Approx. 15'000 - precise number needs to validated	Yes

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

					probenahme-schadstoffuntersuchungen.html. Various publications of the National Soil Monitoring networtk NABO describe the sampling methods in their reports, see publications on https://www.nabo.admin.ch.								
Turkey	2020	point based not all from the soil profile	200	32000	agricultural areas and forest; random and grid method	Yes	WGS 84/UTM Zone	none	200	7743	24057	Yes	
United Kingdom (Northern Ireland)	2020	Point and polygon data held.	Appox 500	Approx 500 (in 5km dataset)	Stratified sample at 5km intersection.	Yes	Irish National Grid - EPSG: 29902	Zero	Approx 50	Approx 450	Zero	The 5K attribute dataset has not been updated, however other dataset have.	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Table 3 (green color for data publicly available, red for not publicly available, light blue for intermediate accessibility, blue-gray for Turkey which has a digital soil database in process of establishing, grey for Spain which does not have a digital database of soil properties)

Country	Year	Does your institution use a common list of soil properties?	Part 1: General Properties	How many profiles have General Properties ?	Part 2: Plant Nutrients	How many profiles have accompanying data on plant nutrients?	Part 3: Soil Salinity	How many profiles have accompanying data on soil salinity?	Part 4: Soil Pollution / Contamination	How many profiles have Soil Pollution / Contamination
Austria1	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Coarse fragments (stoniness), pH, Water storage capacity, horizon term, content of carbonates, structure, mottles, concretions, Munsel color	12000	None	0	Electric conductivity (EC), Carbonate (CO3--)	20	None	0
Austria2	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	9500	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Boron (B), Chlorine (Cl), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	9500	Electric conductivity (EC)	5000	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.), POPs (e.g. pesticide residues, pharmaceuticals, etc)	10000 (heavy metals for nearly all sites, organics and POPs less

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Belgium	2018	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity, Soil biodiversity parameter	0	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Boron (B), Chlorine (Cl), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	0	Electric conductivity (EC), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++), Chloride (Cl-), Sulfate (SO4--), Carbonate (CO3--)	0	None	0
Denmark	2020	Yes	Soil type, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	551	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Copper (Cu), Iron (Fe), Zinc (Zn)	1976	None	0	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co)	12
Finland	2018	No	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	3	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Boron (B), Chlorine (Cl), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	3	Magnesium (Mg++), Calcium (Ca++), Sulfate (SO4--)	1	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co)	1
France	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity)	Not all the profiles have the same informatio	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	Not all the profiles have the same informatio	Electric conductivity (EC), Exchangeable Sodium Percentage	Not all the profiles have the same information. It is far too	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs, PCBs,	Not all the profiles have the same information. It is far too

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				n. It is far too complicated to describe.		n. It is far too complicated to describe.	(ESP), Sodium Adsorption Ratio (SAR), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++)	complicated to describe.	Dioxins/Furans, Chlorinated pesticides	complicated to describe.
Germany 1	2018	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	1800	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Copper (Cu), Iron (Fe), Manganese (Mn), Zinc (Zn)	1800	None	0	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.), POPs (e.g. pesticide residues, pharmaceuticals, etc), NO3-N 1:2 extraction	1800
Germany 2	2018	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, inorganic carbon, stratigraphy, site description, description of soil horizons	3100	Nitrogen (N)	3100	Electric conductivity (EC), Carbonate (CO3--)	3100	None	0
Hungary1	2020	No	Soil depth, Organic carbon / organic matter, pH, CaCO3 content	Soil depth – all locations; SOC ~12,000; pH ~55,000;	None	-	Total Soluble Salts (TSS)	Total Soluble Salts (TSS) ~12,000	None	-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				CaCO3 ~32,000						
Hungary2	2020	No	Soil type, Soil depth, Organic carbon / organic matter, Sand content, Silt content, Clay content, Bulk density (measured), pH, CEC (cation exchange capacity), Water storage capacity	1236	Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Boron (B), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	1236	Exchangeable Sodium Percentage (ESP), Total Soluble Salts (TSS), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++)	1236	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), POPs (e.g. pesticide residues, pharmaceuticals, etc)	Heavy metals: 1236, pesticide residues: 100 (between 1994-1997, and in 2000), organic micro-pollutants: 30-80 sampling places (1996-1997)
Hungary3	2020	No	Organic carbon / organic matter, Bulk density (measured), pH, Soil biodiversity parameters, CaCO3	1959 soil profiles, and 5817 soil samplings	Nitrogen (N), Phosphorus (P), Potassium (K), Magnesium (Mg), Copper (Cu), Manganese (Mn), Zinc (Zn)	1959 soil profiles, and 5817 soil samplings	None	-	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), POPs (e.g. pesticide residues, pharmaceuticals, etc)	57 soil profiles for POPs, 141 soil profiles for heavy metals

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Hungary4	2020	No	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), pH, CEC (cation exchange capacity), Water storage capacity	2129 soil profiles (5976 soil samples)	None	0	Exchangeable Sodium Percentage (ESP), Total Soluble Salts (TSS)	3356 soil profiles with TSS; 1163 profiles with TSS, exchangeable Na+; 9 profiles with TSS, exchangeable Na+, exchangeable Mg++, exchangeable Ca++	None	0
Italy	2018	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	13857	Nitrogen (N)	4817	Electric conductivity (EC)	5845	None	NONE
Latvia1	2020	Yes	Soil type, Organic carbon / organic matter, pH, Soil granulometric composition, relief, stoniness, humidity	Profiles : None	Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Boron (B), Copper (Cu), Iron (Fe), Manganese (Mn), Zinc (Zn), Na	Profiles none: Only for topsoil (20cm depth)	None	None	None	None
Latvia2	2020	Yes	Soil type, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity)	95	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	95	Sodium (Na+), Magnesium (Mg++), Calcium (Ca++), Carbonate (CO3--), Bicarbonate (HCO3-)	95	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co)	95

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 862695

Lithuania	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Coarse fragments (stoniness), pH	107 618	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Iron (Fe)	107 618	Sulfate (SO4--), Carbonate (CO3--)	107 618	None	None
Netherlands	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	Llittle, the first 7 are always determined	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	Not many	Sodium (Na+), Magnesium (Mg++), Calcium (Ca++)	None	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co)	358
Norway	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	Approx. 2000 (not checked)	Nitrogen (N), Phosphorus (P), Iron (Fe)	Approx. 2000 (not checked)	Sodium (Na+), Magnesium (Mg++), Calcium (Ca++)	Approx. 2000 (not checked)	None	None
Poland	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Coarse fragments (stoniness), pH	10000	Phosphorus (P), Potassium (K), Magnesium (Mg)	10000	Carbonate (CO3--)	10000	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co)	50000 topsoil samples

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Slovakia	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity)	18059 (318 also for bulk density)	Nitrogen (N), Phosphorus (P), Potassium (K), Magnesium (Mg), Copper (Cu), Manganese (Mn), Zinc (Zn)	18059 for P and K only, 318 also for all for others	Electric conductivity (EC), Exchangeable Sodium Percentage (ESP), Sodium Adsorption Ratio (SAR), Total Soluble Salts (TSS), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++), Chloride (Cl-), Sulfate (SO4--), Carbonate (CO3--)	8	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), Se, Fe (Fe - only regional)	59 (As - 8, Cd - 5, Co - 14, Cu - 7, Ni -8, Pb - 3, Zn - 6, Hg - 8)
Switzerland	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity, Soil biodiversity parameters, Other: as well as other data, see data model documentation on https://www.nabodat.ch/index.php/de/fachapplikation/datenmodell	Very few: soil biodiversity parameters are rarely included. There are approx. 12'000 fully described soil profiles (except soil biodiv.	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Boron (B), Chlorine (Cl), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	Very few	Electric conductivity (EC)	None	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.), POPs (e.g. pesticide residues, pharmaceuticals, etc)	Very few (some, but not all soil monitoring sites have a full pedological description))

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				parameters))						
Turkey	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), pH	200 soil profiles, 32000 surface samples	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Boron (B), Copper (Cu), Iron (Fe), Manganese (Mn), Zinc (Zn), heavy metals (Cd, Cr, Ni, Pb)	32000 samples	Electric conductivity (EC), Exchangeable Sodium Percentage (ESP), Sodium Adsorption Ratio (SAR), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++)	32000 samples	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), Cd, Cr, Ni, Pb	32000 samples
United Kingdom (Northern Ireland)	2020	Yes	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	All, apart from Bulk density & water storage where only approx 150 profile measurements.	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	All	None	None	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co)	Approx 250

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Table 4 (green color for data publicly available, red for not publicly available, light blue for intermediate accessibility, blue-gray for Turkey which has a digital soil database in process of establishing, grey for Spain which does not have a digital database of soil properties)

Country	Year	What is the format of the database?	What kind of metadata is included in the database?	What metadata standards are used?	link to documentation describing the database structure_	Please describe the database structure (optional)	Do you have a quality control for the data in the database?	Please describe the quality control/assessment procedure	Please indicate main challenges and capacity lacks
Austria1	2020	PostgreSQL	Time (date) of soil survey / soil analysis, Data source		https://bodenkarte.at/#/l/ba,false,60,kb		No		Lack of skilled stuff, financial resources
Austria2	2020	ORACLE	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	BORIS / Data key soil science, INSPIRE			Yes	Data harmonisation is done due to the regulations of the Data Key Soil Science, data import is done following the interface structure of the database, during import automated checks of consistency and plausibility are run	secure funding, harmonisation of methods, data protection, data ownership
Belgium	2018	PostgreSQL	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data	INSPIRE compliant			Yes	quality manual	time and money

			authorship/ownership						
Denmark	2020	MS Access, additional data as Excel spreadsheets	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis	Not known	https://tidsskrift.dk/geografisktidsskrift/article/download/44074/52611	Table of profiles links to table with horizons for each profile, which again links to tables with soil measurements	No		Lack of skilled stuff, Lack of project time for this task
Finland	2018	csv	None		http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/15905/1/lbna24729enc.pdf		Yes	ring test of labs	Lack of communication between organizations
France	2020	Postgresql	postgresql metadata + dublin core + list of codes	dublin core	https://dw3.gissol.fr/fichiers/dictionnaire_donesol_igcs_3-11_01-03-2020.pdf	see the dictionary	yes	data entry interface + automatic and manual controls	Lack of skilled staff, financial resources
Germany1	2018	PostgresSQL	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source	?	https://www.thuenen.de/de/wo/arbeitsbereiche/bodenschutz-und-waldzustand/bodenzustandserhebung/	Thematically organized relational database on PostgreSQL system. Hierarchically structured tables with versioning using views.	Yes	Primary key, unique keys, versioning. Extensive metadata system describing the whole attribute catalog: defined metric attributes, codes using lookup tables. Many check routines, including cross validation.	time

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Germany2	2018	MySQL	None	none	not available		Yes	lab measurements and site description (field) are compared; lab measurements are checked for each individual sample; lab conductes quality checks of measurements/instruments regularly	Lack of national conceptual model and standard, lack of permanent staff
Hungary1	2020	Microsoft SQL, shp	None				No		Lack of national conceptual model and standard
Hungary2	2020	FOXPRO	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis				No		Lack of communication between organizations, Lack of national conceptual model and standard
Hungary3	2020	MySQL	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership				Yes	The projekt developed an own quality control system. E.g. Proper site description was ensured during the survey (geotagged environment photos, photo of completed paper report). Measured	Lack of national conceptual model and standard

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

								values had to fit into a predefined interval.	
Hungary4	2020	RData, csv	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, unit, description of the soil property, reference for measurement methods	Metadata of SoilGrids was considered for providing metadata of MARTHA.	Previous version of the dataset was published in Makó et al. (2010) and Tóth et al. (2012). In the meanwhile the dataset has been restructured and extended with data of further soil profiles, topographical, meteorological, geological information and remot	The MARTHA dataset is structured into GENERAL, ENV_COV, BASIC, CHEMICAL, PHYSICAL,PSD, HYDROLOGICAL, RET, VG_PARAM and METHOD subtables. The subtables can be merged based on PROFILE_ID or SAMPLE_ID.	Yes	Information checked in the database: - checking replicate samples; - taxonomic information: correcting mistyping of Hungarian soil types, checking if different taxonomy levels are in line; - soil depth, soil chemical and physical properties: outliers has	Lack of communication between organizations
Italy	2018	MS ACCESS	Measurement units, Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	ITALIAN STANDARD	http://sito.entecra.it/portale/cra_manuali_dettaglio.php?id_manuale=8915&lingua=IT	RELATIONAL DATABASE	Yes	AUTHOMATIC CHECK OF PERCENTAGE FOR TEXTURE	NONE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Latvia1	2020	Microsoft SQL server	Measurement units, time (date) of soil survey / soil analysis, ownership, other: "in our database all agrochemical research data is well structured - lots of metadata columns are described there			This database is based on GIS. Data processing takes place in the GIS ArcMap software, where this information is later sent to the Crop management information system of Latvia database. The database information is handled by the cartograph and the laborat	Yes	Data validation against the threshold values. Agronome tests the results obtained by the laboratory.	Lack of communication between organizations, Lack of national conceptual model and standard, Funding
Latvia2	2020	open document (ods), dbf / shp	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis	LVS / ISO	ICP Forests, Biosoil demonstration project data stucture, available on demand		Yes	Intercalibration of analytical procedures and sampling	Lack of skilled stuff, Insufficient equipment supply, Lack of funding
Lithuania	2020	dbf, shp	Soil analysis methods, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership				Yes	With the help of ArcGIS program tools and the estimation of soil experts	Lack of skilled stuff

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Netherlands	2020	Oracle, xls	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	National model for the Key Registry, own model for BIS, flat xls for Geochemical atlas	Yes	Internal review procedure by soil scientists, validation of maps is undertaken with independent profile or sample descriptions, quality control program of National Research Tasks support	Several databases present, only partial integration foreseen	Oracle, xls	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership
Norway	2020	PostgreSQL/GIS	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	ISO-standards for the analysis in the lab			Yes	Data validation is done according to code lists and spatial reference rules in the institute	None of them
Poland	2020	shp/dbf, MySQL	Time (date) of soil survey / soil analysis	no specific standards			Yes	examining manually typed data across the database, e.g. removal of pH results out of typical range	databases are well developed but challenges are related to financial resources for database maintenance and update

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Slovakia	2020	MS SQL Spatial, Oracle	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Measurement units	Datasets: ISO 19115; Services: ISO 19119		Some datasets described in: GS Soil project deliverable D4.1: Baritz, R.(ed). 2010. Theme specific test cases for developing data specifications for spatial soil information.	Yes	Code-level check, based on responsible persons according to monitored threats to soil	Lack of skilled stuff, Insufficient equipment supply
Switzerland	2020	Oracle	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership, Data visibility / access restriction parameters, as well as a number of data model hierarchical metainfo (such as info on the owner of the site; project that use this site etc.)	Metadata standards, e.g. as given by the national law of geoinformation and respective documents, are used whenever possible. The metadata is described in the data model. The database structure follows the data model https://www.	The database structure follows the data model https://www.nabodat.ch/index.php/de/fachapplikation/datenmodell. A digital Entity Relationship model is available upon request	The database structure follows the data model https://www.nabodat.ch/index.php/de/fachapplikation/datenmodell. A digital Entity Relationship model is available upon request	Yes	Data Quality is a crucial factor determining applicability and db success. A lot of manpower has been invested to ensure that imported data has undergone rigorous testing, on a formal basis (data type match & value boundaries tests), as well as a scientific, pedological basis through expert judgement. Missing metadata (e.g. sampling methods) have been completed if possible	Lack of skilled stuff, Lack of communication between organizations, Lack of skilled staff

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				nabodat.ch/index.php/de/fachapplikation/datenmodell. A digital Entity Relationship model is available upon request.					
Turkey	2020	Excel	Soil analysis methods, Methods of chemical extraction, Time (date) of soil survey / soil analysis, Data source, Data authorship/ownership	Turkish Stadarts Institution (TSE) Soil Survey Manual			Yes	TSE 17025	Lack of skilled stuff, Lack of communication between organizations, Lack of national conceptual model and standard
United Kingdom (Northern Ireland)	2020	Access	None			Individual soil profiles with soil parameters as fields.	No		Lack of skilled stuff

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Questions related to the Soil Information System

The following questions were made:

Does your country have a Soil Information System (SIS), which includes the database described in the previous sections?
Name and affiliation of the SIS
Reference (website/paper)
Who are the users of the SIS?
For what purposes SIS is used in your country?
Do you have a data infrastructure for data sharing and download (e.g. product catalogue, catalogue service, other web services)? Please specify.
Does your country/institution have a data policy?
Please describe the data policy used for data sharing through the SIS
What components does the SIS have (including the data served from external providers)?
Please specify soil data type and format
Please specify land use / land management data type and format (including the data served from external providers)
Please specify supplementary environmental data type and format (including the data served from external providers)
Please specify soil threats related data type and format (including the data served from external providers)
What are the data sharing protocols? (e.g. WMS, WFS, http, ftp, etc…)
Please, describe the principles of organization of the SIS (optional)
Do you have assessment of soil / land suitability for certain crops and agricultural systems included in the SIS?
Do you have the methods in place to derive specific soil parameters and indicators from measured soil properties (pedotransfer rules)?
Please list the soil-related parameters and indicators derived using the pedotransfer rules
Please indicate main challenges and capacity lacks in developing and maintaining a soil information system

The answers received are reported in the following the tables 5 and 6.

Table 5 (green color for data publicly available, red for not publicly available, light blue for intermediate accessibility, grey for countries who did not describe a SIS in the questionnaire)

Country	Year	Does your country have a Soil Information System (SIS)?	Name and affiliation of the SIS	Reference (website/paper)	Who are the users of the SIS?	For what purposes SIS is used in your country?	Do you have a data infrastructure for data sharing and download	Does your country/institution have a data policy?	Please describe the data policy used for data sharing through the SIS
Austria1	2020	No							
Austria2	2020	Yes	BORIS Soil Information System	www.borsidaten.at	Soil experts/data owners, Public, Universities, scientists, experts in national and federal institutions, experts & specially certified engineers working in the field of Strategic Environmental Assessment (SEA) & Environmental Impact Assessment (EIA)	Strategic Environmental Assessment (SEA) & Environmental Impact Assessment (EIA), national and European soil related projects, reporting for European and national strategies for soil protection, soil assessment for different purposes	two Applications: BORIS EXPERTS (for data owners, showing all data analytical values and Details), BORIS PUBLIC showing just soil site- and profile information, no analytical values. Tools for data selection and presentation & download via webGIS	No	
Belgium	2018	Yes	Bodemdatabank DOV	dov.vlaanderen.be	diverse users	diverse purposes	INSPIRE compliant product catalogue, catalogue service, other web services	No	
Denmark	2020	No							
Finland	2018	No							

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

France	2020	Yes	Système d'information Sol, INRAE	Le système d'information national sur les sols : DONESOL et les outils associés. Grolleau, E., Bargeot, L., Chafchafi, A., Hardy, R., Doux, J., Beaudou, A., Le Martret, H., Lacassin, J.C., Fort, J.L., Falipou, P., Arrouays, D. (2004). Étude et Gestion des Sols, 11 (3), 255-269.	Universities, students, researchers, institutes, Ministries, private companies	Databases are used to provide information about soil distribution, properties and quality. It is also use as input parameters for spatial modelling and analyses (for agricultural and environmental issues).	interactive map in web browser, WMS service, downloadable data	In development	For public data, data are open through institutional portals, other data are accessible by licencing, mainly for research purposes
Germany1	2018	In process of establishing	BZE Wald	https://www.thuenen.de/de/wo/arbeitsbereiche/bodenschutz-und-waldzustand/bodenzustandserhebung/	company, NGO, public, science	INSPIRE, data download	web services (only first inventory	Yes	at the moment: ask via e-mail for data, in future: free agregated data
Germany2	2018	No							
Hungary1	2020	No							
Hungary2	2020	No							
Hungary3	2020	No							

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Hungary4	2020	Yes	DOSoReMI.hu (Digital, Optimized, Soil Related Maps and Information in Hungary), Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research	http://dosoremi.hu	Universities, lecturers, students, researchers, institutes.	The database is used as soil information parameters for spatial modelling and analyses.	interactive map in web browser, WMS service, not downloadable data	No	
Italy	2018	Yes	SOIL INFORMATION SYSTEM OF ITALY, CREA-AA	SOILMAPS.IT	CREA-AA, GENERAL PUBLIC	RESEARCH	YES ON SOILMAPS.IT	Yes	IODL2.0 OPEN SOURCE
Latvia1	2020	No							
Latvia2	2020	No							
Lithuania	2020	Yes	Land information system LIS (ŽIS). LIS supervisor - National Land Service under the Ministry of Agriculture	https://www.geoportal.lt/geoportal/web/en	Land managers, owners and users, students and researchers, etc.	Land management and spacial planning.	htpps://zis.lt/en/paslaugos/	Yes	The data is publicly available
Netherlands	2020	Yes	National Key Registry of the Subsurface (Basis Registratie Ondergrond) and BIS	https://basisregistratieondergrond.nl/english and https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Facilities-Tools/Dutch-	See above	See above	Wms, wfs, viewers	Yes	Governmental daya are open and freely available

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				Soil-Information-System-SIS.htm					
Norway	2020	No							
Poland	2020	Yes	ZSIRPP	http://www.iung.pulawy.pl/images/wyd/pib/zesz51.pdf	regional administration, students, science, agencies	spatial planning, CAP analysis, delineation of regions for support, policy implementation, local analysis	No	No	
Slovakia	2020	Yes	Pôdny portál (EN: Soil Portal)	www.podnemapy.sk	goverment, municipalities, soil users, reserchers, public	for decision and control purposes	no	Yes	Country: http://www.informatizacia.sk/ext_dok-strategicky_dokument_2014_2020_en/16622c
Switzerland	2020	Yes	NABODAT is both the db and a SIS. But there is currently no web-GIS interface to NABODAT	www.nabodat.ch	Please see answer to the same question on the db	Please see answer to the same question on the db	Partly. the publicly available soil data can be requested on https://www.nabodat.ch/index.php/de/service/bodendatensatz	Yes	The data policy is open according to Aarhus and Swiss Open Governmental Data principles. I think this is corresponding to the https://creativecommons.org/licenses/by-sa/4.0/ licence, but this is not verified by our legal department. There is also a subset of the data that is considered to be personal data which is more strongly

									protected and less accessible
Turkey	2020	No							
United Kingdom (Northern Ireland)	2020	No							

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Table 6 (green color for data publicly available, red for not publicly available, light blue for intermediate accessibility)

Country	Year	What components does the SIS have?	Please specify soil data type and format	Please specify land use / land management data type and format	Please specify supplementary environmental data type and format	Please specify soil threats related data type and format	What are the data sharing protocols? (eg WMS, WFS, http, ftp)	Please, describe the principles of organization of the SIS	Assessment of soil / land suitability	Methods to derive specific soil parameters and indicator	Soil-related parameters and indicators	Please indicate main challenges and capacity lacks for SIS
Austria2	2020	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples	None	None	Point data - soil monitoring sites	WMS / WFS		No	No		legal implementation of soil surveys and soil information system, strongly connected to implementation of soil protection laws & regulations, raising new data - legal implementation of soil surveys

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Belgium	2018	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Polygon data - soil mapping units, Polygon data - artificial borders (e.g. agricultural parcels) with soil data, Raster data - derived	Polygon/raster data - land use types	Polygon/raster data - topography (e.g. elevation, slope, etc.), Polygon/raster data - geology (e.g. geology map, map of Quaternary sediments, etc.), Polygon/raster data - water (e.g. watersheds, water quality for irrigation, ground water level, etc.)	Point data - soil monitoring sites, Raster data - carbon sequestration potential, Polygon/raster data - soil erosion risk, Polygon/raster data - salt-affected areas, Polygon/raster data - soil sealing	WMS, WFS, http		No	No		time and money
France	2020	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, soil analyses, Point data - topsoil samples, Polygon data - soil mapping units, soil typological units, layers, Polygon data - artificial borders (e.g. agricultural parcels) with soil	for soil profiles : nature of vegetation, nature of land use (completion variable); for soil mapping units : nature of land use ; for soil monitoring data : land management data on fertilization, agricultural operation dates including pesticides application (completion variable) - from external providers : corine land cover, ecoclimap land use	Raster data - Digital elevation model, Polygon/raster data - geology (e.g. geology map, map of Quaternary sediments, etc.)	Data on soil erosion risk from modelling	WMS, WFS, http	Relational database: data model explained in the DoneSol data dictionary	donesolweb interface for soil data entry, donesol database collecting soil data on soil profiles, soil horizons, soil analyses, soil mapping and soil typological units (data organized by pedological studies), other specific databases, datawarehouse with statistical soil information	Yes	Bulk density, P olsen, Soil Water Content, Aqua Regio/ HF, soil available water, parameters for modelling (depending on the databases to which rules are applied)	Lack of skilled staff, financial resources

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

			data, Raster data - derived soil properties and indicators, Polygon data - derived soil properties and indicators	data, statistical land management data					system, database for automatic quality control storing SQL queries			
Germany1	201 8	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Raster data - derived soil properties and indicators	Polygon/raster data - land use types, Polygon/raster data - major crops, Polygon/raster data - agricultural systems	Polygon/raster data - land cover types (e.g. cropland, grassland, forest, urban, etc.), Polygon/raster data - topography (e.g. elevation, slope, etc.), Polygon/raster data - climate (e.g. mean temperature, rainfall, etc.), Polygon/raster data - geology (e	Point data - soil monitoring sites, Point data - contaminated sites, Raster data - carbon sequestration potential, Polygon/raster data - soil acidification, nutrition deficiency	WFS, data downloading		forest types, vegetation type	Yes	Field capacity, useable field capacity	time
Hungary4	202 0	Soil data, Topographic or satellite imagery base map.	Raster data - derived soil properties and indicators	None	None	None	WMS		No	Yes	hydraulic pedotransfer functions	Lack of communication between organizations

Italy	2018	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Polygon data - soil mapping units, Polygon data - artificial borders (e.g. agricultural parcels) with soil data, Raster data - derived	Polygon/raster data - land use types	Polygon/raster data - climate (e.g. mean temperature, rainfall, etc.)	Raster data - carbon sequestration potential	NOT WORKING AT THE MOMENT	SITE-HORIZONS-TYPOLOGICAL UNITS (DERIVED SOIL PROFILE)-SOIL POLYGONS AT VARIOUS SCALES	No	Yes	bulk density, available water capacity, soil moisture control section, cation exchange capacity, depuration capacity, protective capacity, saturated hydraulic conductivity, land capability, hydrologic soil group, pedoclimate, crusting risk, compaction ris	NONE
Lithuania	2020	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Administrative boundaries, Spatial datasets of: (1) reclamation status and sodden soils; (2) limited land use areas; (3)	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Polygon data - soil mapping units	Polygon/raster data - land use types, Polygon/raster data - major crops, Polygon/raster data - irrigation, Polygon data - cadastral information, Polygon data - land (soil) productivity evaluation data (in points).	Polygon/raster data - land cover types (e.g. cropland, grassland, forest, urban, etc.), Polygon/raster data - topography (e.g. elevation, slope, etc.)	Point data - soil monitoring sites, Polygon/raster data - eroded soils	dbf; shp		Yes	No		Lack of skilled stuff

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Netherlands	202 0	Soil data	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Polygon data - soil mapping units, Raster data - derived soil properties and indicators, Polygon data - derived soil properties and indicators	None, this is included in other Key Registries or information systems	Polygon/raster data - geology (e.g. geology map, map of Quaternary sediments, etc.)	Soil compaction risk	Wms, wfs		No	Yes	Soil hydrophysical properties such as bulk density, water holding capacity, infiltration capacity etc	See above
Poland	202 0	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Polygon data - soil mapping units, Polygon data - derived soil properties and indicators	Polygon/raster data - land use types	Polygon/raster data - land cover types (e.g. cropland, grassland, forest, urban, etc.), Polygon/raster data - topography (e.g. elevation, slope, etc.), Polygon/raster data - climate (e.g. mean temperature, rainfall, etc.)	Point data - soil monitoring sites, Point data - contaminated sites, Polygon/raster data - contaminated areas, Polygon/raster data - soil erosion risk, Polygon/raster data - soil acidification	WMS for limited data		Yes	Yes	water retention, susceptibility to erosion, susceptibility to compaction	financial resources for maintenance and update

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Slovakia	202 0	Soil data, Land use / land management data, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Polygon data - soil mapping units	Polygon/raster data - major crops	Polygon/raster data - land cover types (e.g. cropland, grassland, forest, urban, etc.), Polygon/raster data - topography (e.g. elevation, slope, etc.)	Point data - soil monitoring sites, Point data - contaminated sites, Polygon/raster data - contaminated areas, Polygon/raster data - soil erosion risk, Polygon/raster data - salt-affected areas, Polygon/raster data - soil compaction, Polygon/raster data -	WMS, WFS, ftp, SOAP, REST		Yes	Yes	bulk density, available soil water capacity, soil compaction	Lack of skilled staff, Insufficient equipment supply
Switzerland	202 0	Soil data, Land use / land management data, Supplementary environmental data (land cover, topography, climate, geology, water, etc),, Soil threats related data (soil erosion, contamination, acidification, salinization, etc.), Administrative boundaries	Point data - soil profiles, Soil horizons (layers) data associated with profile location, Point data - topsoil samples, Polygon data - soil mapping units, Polygon data - derived soil properties and indicators	None	None	Point data - soil monitoring sites, Point data - contaminated sites, The focus of NABODAT is on Soil pedological description and soil pollution. Other threats are rarely	The formerly offered WFS has been terminated due to its huge protocol overhead and subsequent extremely large downloads. Data can be downloaded as Excel	NABODAT is a web application that can be accessed from anywhere (require access to internet). The main navigation GUI is depicted and described in http://www.nabodat.ch/Anwenderhandbuch. The navigation follows primarily the data model hierarchy	No	No		Lack of skilled stuff, Lack of communication between organizations, Lack of skilled staff

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

						stored in the db	sheets, in the INTERLIS export format or as file-based geodatabase					

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Questions related to the Soil Monitoring System

The following questions were made:

Does your country have a digital soil monitoring system?
Which soil degradation processes (soil threats) are being monitored?
Part 1: General Properties
Part 2: Plant Nutrients
Part 3: Soil Salinity / Alkalinity
Part 4: Soil Pollution / Contamination
Please indicate main challenges and capacity lacks in developing and maintaining a soil monitoring system
Do you think it is necessary for your country to develop a SIS?
Comments / remarks

The answers received are reported in the table 7.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Table 7

Country	Year	Does your country have a digital soil monitoring system?	Which soil degradation processes are being monitored?	Part 1: General Properties	Part 2: Plant Nutrients	Part 3: Soil Salinity / Alkalinity	Part 4: Soil Pollution / Contamination	Please indicate main challenges and capacity lacks in MONITORING	Do you think it is necessary for your country to develop a SIS?	Comments / remarks
Austria1	2020	Yes	Soil monitoring is not task of the soil map, but occasion related of map content upgrades however (e.g. water conditions, Corg-content)	None	None	None	None	Lack of skilled stuff, financial ressources	Yes	
Austria2	2020	No							Yes	Important is to ensure the service of existing tools and soil information systems for the future, to enforce the regular raising of new data for known and new sites and new parameters (monitoring) due to current soil protection issues (e.g. (micro-)plast

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Belgium	2018	In process of establishing	Soil erosion, Loss of soil organic matter, Soil sealing	Organic carbon / organic matter content, Organic carbon stocks, texture, bulk density	None	None	None	money	Yes	We have a historical database of the profiles and soil samples of the soil mapping campaign of Belgium. We are developing a new soil database that can centralize very diverse soil data from more recent soil sampling and profile descriptions from Flanders.
Denmark	2020	Yes	Loss of soil organic matter, Nutrient imbalance	Organic carbon / organic matter content, Organic carbon stocks	Nitrogen (N), Phosphorus (P)	None	None	Not known	Yes	
Finland	2018	No							Yes	
France	2020	Yes	-Decline in Soil organic C - Acidification - Decline in soil fertility - Soil contamination - Decline in soil biodiversity	Soil type, Soil depth, Organic carbon / organic matter, Texture class, Sand content, Silt content, Clay content, Bulk	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum	Electric conductivity (EC), Exchangeable Sodium Percentage (ESP), Sodium Adsorption Ratio (SAR), Sodium (Na+), Magnesium	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs, PCBs, Dioxins/Furans, Chlorinated pesticides	Money issues !		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				density (measured), Coarse fragments (stoniness), pH, CEC (cation exchange capacity), Water storage capacity	(Mo), Zinc (Zn)	(Mg++), Calcium (Ca++)				
Germany1	2018	Yes	Loss of soil organic matter, Nutrient imbalance, Soil acidification, Loss of biodiversity, Soil pollution / contamination	Organic carbon / organic matter content, Organic carbon stocks, pH, Soil biodiversity parameters	Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Copper (Cu), Iron (Fe), Manganese (Mn), Zinc (Zn)	None	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.), POPs (e.g. pesticide residues, pharmaceuticals, etc)	we have a monitoring system	No	INSPIRE could be discussed how handle different countries
Germany2	2018	In process of establishing	Loss of soil organic matter	Organic carbon / organic matter content, Organic carbon	Nitrogen (N)	Electric conductivity (EC), Carbonate (CO3--)	None	Lack of national conceptual model and standard, lack of permanent staff	Yes	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

				stocks, pH, bulk density						
Hungary1	2020	No							Yes	
Hungary2	2020	Yes	Loss of soil organic matter, Soil acidification, Soil pollution / contamination, Salinization / sodification	Organic carbon / organic matter content, pH, CaCO3 content, Total N content	Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Boron (B), Copper (Cu), Iron (Fe), Manganese (Mn), Molybdenum (Mo), Zinc (Zn)	Exchangeable Sodium Percentage (ESP), Total Soluble Salts (TSS), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++)	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), pesticide residues: between 1994-1997, and in 2000; organic micro-pollutants: 1996-1997	Lack of communication between organizations, Lack of national conceptual model and standard, ongoing measurements are limited because of financial causes	Yes	
Hungary3	2020	No								
Hungary4	2020	Yes								
Italy	2018	No								
Latvia1	2020	Yes	Organic carbon/organic matter, pH	Organic carbon / organic matter content, pH	Phosphorus (P), Potassium (K), Magnesium (Mg)	None	None	Lack of communication between organizations, Lack of national conceptual model and	Yes	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

								standard, Funding		
Latvia2	2020	No							Yes	
Lithuania	2020	Yes	Soil erosion, Nutrient imbalance	Organic carbon / organic matter content, pH	Nitrogen (N), Phosphorus (P), Potassium (K)	None	None	Lack of skilled stuff	No	
Netherlands	2020	Yes	Monitoring is on soil properties, functions and threats are derived afterwards	Organic carbon / organic matter content, Organic carbon stocks, pH, see other surveys	Nitrogen (N), Phosphorus (P), Potassium (K)	Electric conductivity (EC)	None	the need of monitoring campaigns is decided by policy makers	No	Soil monitoring systems details in other surveys
Norway	2020	No							Yes	The soil information system is a mapping system widely used. The soil property database is merely a soil property reference system used in the soil information system. We have a system capable of monitoring soil changes in soil polygons using a national and statistically valid sampling grid. The link below provides information on the soil information system: https://kartkatalog.geonorge.no/metadata/soil/c961484d-9d4b-4f6a-8bd5-e145e96d1560

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Poland	2020	Yes	Loss of soil organic matter, Nutrient imbalance, Soil acidification, Soil pollution / contamination, Salinization / sodification	Organic carbon / organic matter content, pH	Nitrogen (N), Phosphorus (P), Potassium (K), Magnesium (Mg), Sulfur (S)	Electric conductivity (EC), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++), Carbonate (CO3--)	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.)	financial resources to extend the number of sampling locations	Yes	integration of various data and enlargement of monitoring program
Slovakia	2020	Yes	Soil erosion, Loss of soil organic matter, Nutrient imbalance, Soil acidification, Soil pollution / contamination, Salinization / sodification, Soil compaction, soils used for energetic crops, peatland	Organic carbon / organic matter content, Organic carbon stocks, pH	Nitrogen (N), Phosphorus (P), Potassium (K), Magnesium (Mg), Copper (Cu), Manganese (Mn), Zinc (Zn)	Electric conductivity (EC), Exchangeable Sodium Percentage (ESP), Sodium Adsorption Ratio (SAR), Total Soluble Salts (TSS), Sodium (Na+), Magnesium (Mg++), Calcium (Ca++), Chloride (Cl-), Sulfate (SO4--), Carbonate (CO3--)	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.)	Lack of skilled stuff, Insufficient equipment supply	Yes	Digital database of agricultural soils of Slovakia described in details above consists of these individual datasets: Digital database of soil profiles, Digital map of land-evaluation units in 1.5k scale, Digital map of soils of Slovakia in 1.400k scale

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695

Switzerland	2021	Yes	Loss of soil organic matter, Soil acidification, Soil pollution / contamination	Organic carbon / organic matter content, Organic carbon stocks, pH, Soil biodiversity parameters	Nitrogen (N)	None	Heavy metals (Pb, Cr, Zn, As, Cu, Hg, Ni, Cd, Co), PAHs (e.g. oil pollution etc.), POPs (e.g. pesticide residues, pharmaceuticals, etc)	Lack of skilled stuff, Lack of communication between organizations, Lack of skilled staff	Yes	This survey has been filled out by fabio.wegmann@bafu.admin.ch on 2021, April 23rd, based on my best currently available knowledge. Please contact me in case of questions or for feedback. I suggest to use a document-based approach for further surveys, as there is no possibility to store intermediate states of filling out the survey. I'd also like to obtain my answers for further reference and discussions. Kind regards, Fabio
Turkey	2020	No							Yes	
United Kingdom (Northern Ireland)	2020	Yes	Loss of soil organic matter, Nutrient imbalance, Soil acidification, Soil compaction	Organic carbon / organic matter content, Organic carbon stocks, pH	Nitrogen (N), Phosphorus (P), Potassium (K), Magnesium (Mg), Sulfur (S)	None	None	Lack of skilled stuff	Yes	The answers here relate only to Northern Ireland. In the wider UK, Cranfield University host the LansIS Soil Information System which contains detailed information on the soils of England & Wales, while the James Hutton Institute host the Soil Information

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 862695