5
GEOSPATIAL STANDARDS
An example from agriculture

Didier G. Leibovici, Roberto Santos, Gobe Hobona, Suchith Anand, Kiringai Kamau, Karel Charvat, Ben Schaap and Mike Jackson

As a preamble, it is worth remembering that geospatial standards relate to: (a) datasets that have geographical location of some sort which can be often represented as a cartographical map or at least being delineated or attached to a geographical area; (b) a set of rules derived from computing science modelling and domain specific science characteristics; and, eventually, (c) processing capabilities allowing to transform and then display the results as a summary, a map, and so on. The definition of a geospatial standard for some data or a service delivery allows data exchange from machine to machine into what is called data interoperability, i.e., understanding the data to be able to operate on it. Different levels of interoperability are often distinguished: syntactic, semantic, structural interoperability. Syntactic interoperability (e.g., format, encoding, protocol and standard) ensures communication and information exchange; structural interoperability (e.g., resolution, spatial accuracy, temporal accuracy, other structural quality, orthorectification) ensures the expected intrinsic characteristic of the data; and semantic interoperability (e.g., classification, measurement attributes, quality of the data, ontology of the data, reference of the information model) ensures an understanding of the meaning of the data and its properties, i.e., enabling users to grasp the knowledge behind the information shared. Thus, different types or levels of standardization may be considered within a particular geospatial standard focusing more often on the syntactic and structural interoperability, with the semantic interoperability adoption coming often from combining multiple viewpoints within a best practice recognized standard and a data specification. Therefore, the level of adoption of standards can vary from one domain to another due to different practices within communities and the range of stakeholders involved.

This chapter is not an attempt to review the evolution of geospatial data standardization in the last 25 years, facilitated since the creation of the Open Geospatial Consortium (OGC) in 1994 and of the ISO/TC211 Technical Committee on geographical information of the International Organization for Standardization (ISO), which was established the same year. It aims to highlight the impact geospatial standards have had so far, as much on their difficult task of harmonizing data exchange between sometimes highly developed disciplinary communities, as on the innovation they enabled. The chapter examines their remits in the current world scenario where the combination of increasing access to high-speed Internet and pervasive computing generates vast amounts of structured and unstructured geographically associated data. These advances are boosting the development of standards, but also expectations in

 DOI: 10.4324/9780367855765-7

relation to data harmonization as well as best practices for a large range of requirements from applications using geolocated information and geospatial data gathered from various sources.

Three main challenges have always been the drivers for innovation in OGC standards development, conditioning their adoption: (1) cross-disciplinary integration; (2) compatibility between national, regional and international e-infrastructures; and (3) efforts required to make possible data conflation from heterogeneous sources including crowdsourcing. Because of the technological advances listed above, these challenges are pressing more and more as the possibilities appear endless. They exist within lots of thematic domains working with legacy geographic information, surveyed data, remote sensing imagery and population data, along with real-time data coming from a range of sensors, Internet of Things (IoT), big data derived from social media, including crowdsourcing, citizen science and volunteered geographic information. From its multiple facets, the agriculture domain is particularly sensitive to the advances in all these respective aspects, from the above challenges to the use and re-use of heterogeneous data within e-infrastructures that enable data transformation and computational models within a range of applications.

The discussion here reflects activities in the last two years within the Agriculture Working Group of OGC, where different domain experts discussed the needs for and from agriculture standards. These standardization efforts are particularly viewed from current and previous EU-funded projects such as agriXchange, FOODIE, DataBio, Demeter and SIEUSOIL. A few case studies involving agriculture illustrate the many issues that have been raised, resolved or still to be resolved or proposing best practices to make the most of their successful adoption.

Besides traditional bodies such as the Food and Agriculture Organization (FAO), the Consultative Group on International Agricultural Research (CGIAR), the European Federation for Information Technologies in Agriculture (EFITA) and the International Society for Precision Agriculture (ICPA), there is also the Global Open Data for Agriculture and Nutrition (GODAN). This is concerned with decision making that is linked to food security and agriculture sustainability from using agriculture information. There are also initiatives from communities of researchers and industrial bodies such as Research Data Alliance (RDA) working groups and the OGC Agriculture Domain Working Group, which undertake initiatives and goals toward the adoption of data and processing standards frameworks for the benefit of all. Other key players include the World Health Organization (WHO), the World Bank and the United Nations committee of experts on Global Geospatial Information Management (UN-GGIM), who are also shaping initiatives for better adoption of developed standards. The Infrastructure of Spatial Information in Europe (INSPIRE), the Group on Earth Observation System of Systems (GEOSS) and the African GEOSS (AfriGEOSS) are examples of concrete outcomes of these efforts.

After the brief historical notes and generic aspects of geospatial standards, the chapter illustrates the process of designing, developing and implementing a standard. Then, it shows how the adoption of geospatial standards plays a role in cross-disciplinary innovation. A section is dedicated to the issues arising from different national policies and international initiatives for the adoption of standards. New paradigms coming from citizens' involvement are analysed in relation to the additional disruptive data flow. Finally, the chapter explores what the future brings as new research challenges in light of the recent developments in the agriculture and related domains.

Historical background

Questions about the standardization of geospatial information started with the growing success of Geographic Information Systems (GIS) in the 1980s within vendor-locked standards

and software. With the advent of Open Source and Open Software, particularly the Geographic Resources Analysis Support System (GRASS), the Open GIS project to develop the vision of data interoperability became the Open GIS Consortium in 1994 and is now the Open Geospatial Consortium. This openness was primordial for the large adoption that we see today and is still the key motivation for its spread (Coetzee *et al*., 2020).

Many of the standards that are implemented across the geospatial industry have been developed and published by the OGC. The OGC is an international consortium of more than 520 businesses, government agencies, research organizations and universities, driven to make geospatial (location) information and services FAIR (Findable, Accessible, Interoperable, and Reusable). OGC runs a member-driven consensus process that creates publicly available, royalty-free, open geospatial standards.

An OGC standard is a document, established by consensus and approved by the OGC membership, that provides rules and guidelines that aim to optimize the degree of interoperability in a given context. The rules and guidelines are specified to take into consideration community and member requirements, as well as market and technology trends. OGC standards endeavour to improve geospatial interoperability across data, processes, workflows, visualization, products, representation and other aspects of geospatial information management.

The OGC is organized into Standards Working Groups (SWGs) and Domain Working Groups (DWGs). SWGs focus on the development of specifications, whereas DWGs provide a platform for discussing topics of relevance to specific communities of interest. The Agriculture DWG is an example of a DWG within the OGC. The Agriculture DWG was formed with a focus on technology and associated policy issues and other technological interests that relate to agriculture as well as the means by which those issues can be appropriately factored into the OGC standards development process. The goal of the group is to establish a link between experts in the domains of agriculture and technology.

The OGC maintains a close formal relationship with ISO/TC211. This relationship has allowed multiple OGC standards to be adopted as ISO standards. For example, the OGC Geography Markup Language (GML) standard was adopted by ISO as ISO 19136:2007 *Geographic information – Geography Markup Language (GML)* – and the OGC Web Feature Service (WFS) standard was adopted by ISO as ISO 19142:2010 *Geographic information – Web Feature Service*. The OGC also maintains a close formal relationship with the World Wide Web Consortium (W3C). This relationship has led to the joint development of multiple specifications, such as the Time Ontology in OWL, Semantic Sensor Network Ontology and Spatial Data on the Web Best Practices specifications (OGC and W3C, 2017a, 2017b). These specifications are maintained by the Spatial Data on the Web Interest Group – a joint working group of the OGC and W3C.

Designing, developing and implementing standards

Geospatial data are applied in multiple domains to improve the understanding of location characteristics and their impact on various aspects of society. Beyond geospatial technologies to capture data, GIS makes it possible to gather data and relate information, within a planned analysis, on real-world phenomena via their location on Earth. From the data capture to the decision making arising after analysis within a GIS, standards are there to seamlessly and efficiently work through a workflow representing the planned analysis. Within a particular domain, if the FAIR principle (described above) is widely adopted, the latter workflow is also effective in cross-domain workflows. A non-exhaustive list of domains that make use of geospatial technologies, data and standards includes: agriculture, aviation, built environment, citizen science, data science, defence, Earth systems science, emergency services, energy and utilities, geology and geophysical

sciences, hydrology, insurance and financial services, land administration, law enforcement, marine, meteorology, simulation and gaming and transportation. This section focuses on soil data, used in a range of applications involving different disciplines and communities in agriculture, as an example of the path of development of a standard from a geospatial data specification.

The endeavour of data interoperability behind development of standards can be very challenging, depending on the domain. Structural and syntactic interoperability consensus may depend on the success of existing data exchange format and standards within a discipline, a community, or directly from the data collection methods (e.g., data capturing devices' capabilities). However, overcoming the semantic hurdle in order to reach domain interoperability is a more difficult task. In other words, legacy data formats and best practices within a domain often incline toward a specific standard framework, but different countries or slightly different communities utilizing the same thematic data may use competitive semantic or ontological frameworks that need converging in order to reach interoperability. Note that this convergence for data interoperability only means to reach compatible standards (i.e., interoperability), not necessarily unifying them. Developments of the European INSPIRE thematic data specifications were possible thanks to the INSPIRE framework and methodology (e.g., working groups, experts, and so on), but with variable progress depending on the thematic data. Reaching data harmonization and FAIRness is conditional to the rate of adoption of the standards and data specifications, and gives a measure of success.

Soil data exchange is a good example of this sometimes long and multifaceted process. In 2010, a markup language for soil data called SoilML was proposed (Montanarella *et al.*, 2010). Pourabdollah *et al.* (2012) integrated the various soil databases approaches of Soil and Terrain (SOTER) Databases, the Food and Agricultural Organization (FAO) classification of 1988 and World Reference Base of Soil Resources (WRB) into the SOTER Markup Language (SoTerML) allowing a flexible characterization and proposed some mapping between the different classifications. SoTerML was designed to be compatible with the Geography Markup Language (GML) standard, an OGC standard built on the Geoscience Markup Language (GeoSciML) for geological data. The INSPIRE soil data specification was established in 2013 and in the same year the Global Soil Map Markup Language (GSMML) was developed (Wilson *et al.*, 2013). Also, in 2013, the ISO published the ISO 28258:2013 standard for the digital exchange of soil-related data. ISO 28258:2013 aims to facilitate the exchange of soil-related data via digital systems between entities such as individuals and organizations. ISO 28258:2013 also uses the GML standard to represent features and to explicitly geo-reference soil data, thus facilitating the use of soil data within GIS. In 2016, the OGC Soil Data Interoperability Experiment was conducted to illustrate the range of different data modelling solutions that are in use, including the SoTerML and the INSPIRE approach (Ritchie, 2016). This OGC Soil Data Interoperability Experiment evaluated existing models and proposed a common core model, including a schema, based on GML, which was tested through the deployment of OGC Web services and demonstration clients.

Hoffman *et al.* (2019) provide a survey of the various standardization activities relating to soil and agriculture. Their study highlights that even with the existence of standards to support soil science and agriculture, there continue to be differences between models adopted by various communities of interest (Batjes *et al.*, 2018; Kempen *et al.*, 2019). To improve interoperability across the different communities, there is a need for joint activities between:

- OGC Agriculture Domain Working Group;
- FAO (Global Soil partnership, i.e., pillar 4 with GLOSIS as the e-infrastructure (see http://www.fao.org/global-soil-partnership/pillars-action/4-information-data/glosis/en/);

- FAO pillar 5 for the harmonization aspects including the soil profile data exchange standard);
- ISO; and
- International Union of Soil Sciences Working Group – Soil Information Standards (IUSS WG-SIS).

Note here the large adoption of XML format (evidenced by use in GeoSciML, INSPIRE and SoilML) as a means of serializing data that are structured according to the data models defined by the data specifications. This is the case for soil data and for a range of environmental data used in agriculture. However, other serialization syntactic formats may also be used to distribute geospatial information over the Internet (e.g., RDF, JSON). Even if in theory one can switch between serialization formats, different modelling communities prefer one or another, which generally facilitates their adoption and evolution.

Emergence of cross-disciplinary enablers

In subdomains of agriculture such as agricultural economics, arable farming, cattle farming, erosion control, mixed farming, soil fertility management, and soil improvement, the geospatial component plays a key role in associating attribute information to a location on the Earth. Such attribute information may include, for example, genetic data, genomics and soil data. Whilst the attribute information may have been collected from a field or farm, in other cases it may be the result of scientific modelling and a simulation, such as in the case of plant growth modelling. Developing data services and e-infrastructure enabling each partner of a multidisciplinary project to visualize data sources from the others is often a driver for data integration generating motivations within teams (Helbig *et al.*, 2017). As a result, strategies and analytics for the visual exploration of complex environmental data are the companions of cross-disciplinary innovation. The geospatial domain structuring of the required attribute information, along with choice of appropriate spatio-temporal scales, allows partners to envisage unforeseen coherence between such multidisciplinary information. These roles also facilitate learning and teaching in agriculture and related environmental studies (Guan *et al.*, 2012). Standards developed and published by the OGC that have been demonstrated to be effective enablers of cross-disciplinary agricultural information exchange include:

- OGC Sensor Observation Service (SOS): This standard specifies a Web service interface for requesting, filtering and retrieving observations and sensor system information. This is the intermediary between a client and an observation repository or near real-time sensor channel.
- OGC SensorThings API (STA): This standard provides an open, RESTful, geospatial-enabled, JSON-based and unified way to interconnect and task IoT devices, data and applications over the Web.
- OGC Sensor Model Language (SensorML): This standard defines models and schema for describing sensor systems and processes and provides information needed for discovery of sensors, location of sensor observations, processing of low-level sensor observations and listing of task-able properties.
- OGC Observations & Measurements (O&M): This standard defines conceptual models for encoding observations and measurements from a sensor, both archived and real-time. Adopted by the International Organization for Standardization (ISO) as ISO 19156:2011.

- OGC Sensor Planning Service (SPS): This standard specifies a web service interface for requesting user-driven acquisitions and observations. This standard allows for tasking sensors.
- OGC Geography Markup Language (GML): The standard specifies an Extensible Markup Language (XML) grammar for expressing geographical features. GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.
- OGC Web Map Service (WMS): This standard specifies an interface that provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases.
- OGC Web Coverage Service (WCS): This standard specifies an interface that offers multidimensional coverage data for access over the Internet.
- OGC Web Feature Service (WFS): This standard specifies an interface that offers direct fine-grained access to geographic information at the feature and feature property level, through discovery operations, query operations, locking operations, transaction operations and operations to manage stored, parameterized query expressions.
- OGC Web Processing Service (WPS): This standard specifies an interface that provides rules for standardizing how inputs and outputs (requests and responses) for geospatial processing, analysis and computational services should be represented.
- OGC Catalogue Services for Web (CSW): This standard specifies an interface that supports the ability to publish and search collections of descriptive information (metadata) for data, services and related information objects.
- OGC WaterML: This standard enables the representation of water observations data, with the intent of allowing the exchange of such datasets across information systems.

To explore and demonstrate how OGC standards can be applied to various domains, the OGC executes testbed projects involving several different governmental, private and academic organizations. Phase 13 of the OGC Testbed series (Testbed-13) included an experiment to test whether some of the standards listed above could enable an application to offer agricultural yield prediction within a Spatial Data Infrastructure (SDI). The experiment and its outcomes are presented by Schumann (2018).

Agricultural yield prediction relies on a variety of factors, such as crop type, disease, relief, soil and weather. Expert knowledge is required to enable the configuration of the variables and sub-variables involved in crop-yield prediction. To enable the integration of an agricultural yield prediction capability within an SDI, participants of Testbed-13 created a wrapper around the Agricultural Production Systems Simulator (APSIM) product using an implementation of the OGC WPS standard. This standard is particularly suited to interfaces for prediction models as it allows for specification of any number of input and output parameters for a process. The participants also created a workflow, using FME software, for executing the WPS and visualizing the results (Figure 5.1).

Amongst the scenarios explored by the testbed was an agricultural scientist client scenario that demonstrated how geospatial standards could enable climate and agricultural science by facilitating the ingest of geospatial data into crop forecasting models, potentially including temporal and spatial subsets of climate model variables relevant to rain and soil conditions. The focus was on providing access to data from the MERRA-2 mission of the National Aeronautics and Space Administration (NASA), exposed through implementation of the OGC WCS 2.0 standard. Moreover, the client application within the scenario also supported the integration of on-demand crop yield prediction outputs from a WPS implementation that

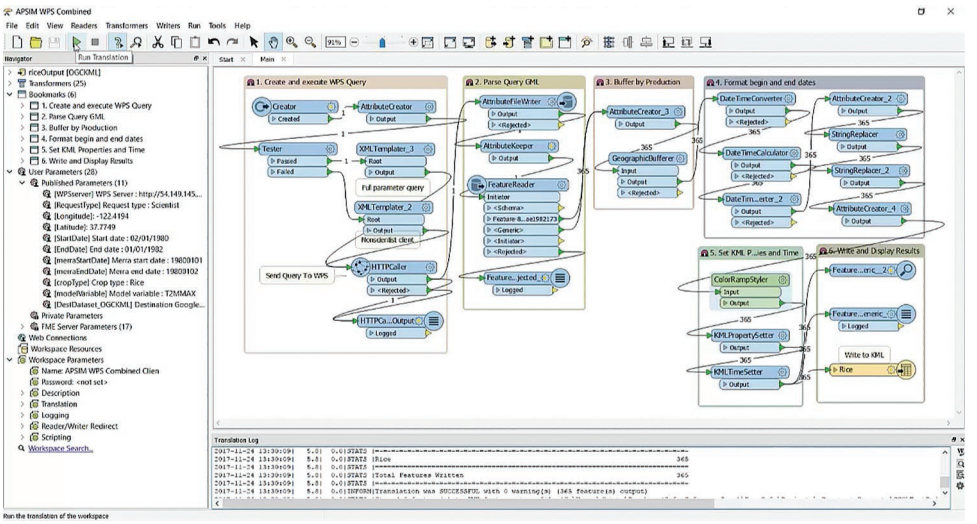

Figure 5.1 APSIM WPS FME combined scientist and non-scientist client implemented using FME workspace

Source: OGC.

had been enabled to provide predictions through running the APSIM crop model. FME was configured to invoke the WPS from the perspective of a scientist user or optionally from the perspective of a non-scientist user. The scientist option employed the full parameter query, whereas the non-scientist mode employed the abbreviated, non-expert query. In both cases, when the client received the WPS response, it parsed the GML contained in the response.

Testbed-13 also examined how a common infrastructure that supports both scientist and non-scientist users could be provided. Using a combination of the Business Process Modelling and Notation (BPMN) and the WPS standards, the testbed participants were able to implement the workflow shown in Figure 5.2.

The workflow enacts a simplistic process for the scientist user that translates the input variables into an APSIM execution document, executes the process and then returns the result. A file containing meteorological information is specified by a scientist within the scientist process. In contrast, for the non-scientist user, the process executes using a few parameters and makes decisions on the type of template APSIM file to use from the crop keyword (a restricted WPS string input). The most suitable file is identified by the workflow engine within the non-scientist process and then augmented with MERRA-2 data for a provided date range.

The architecture for this common infrastructure implemented in Testbed-13 included WCS and WPS implementations. The architecture implemented in the testbed could be extended to include implementations of WFS, SOS and STA to allow for ingesting meteorological information from near-real time observations and forecasts. Further, the architecture could be extended to support geospatial formats such as the OGC GeoPackage standard, which provides a standard way for encoding geospatial concepts in an SQLite database. Both GML and GeoPackage standards support fundamental geospatial concepts such as Coordinate Reference Systems and Simple Feature geometries. These geospatial concepts make it possible to geographically integrate cross-disciplinary data from climate science, agriculture and other fields.

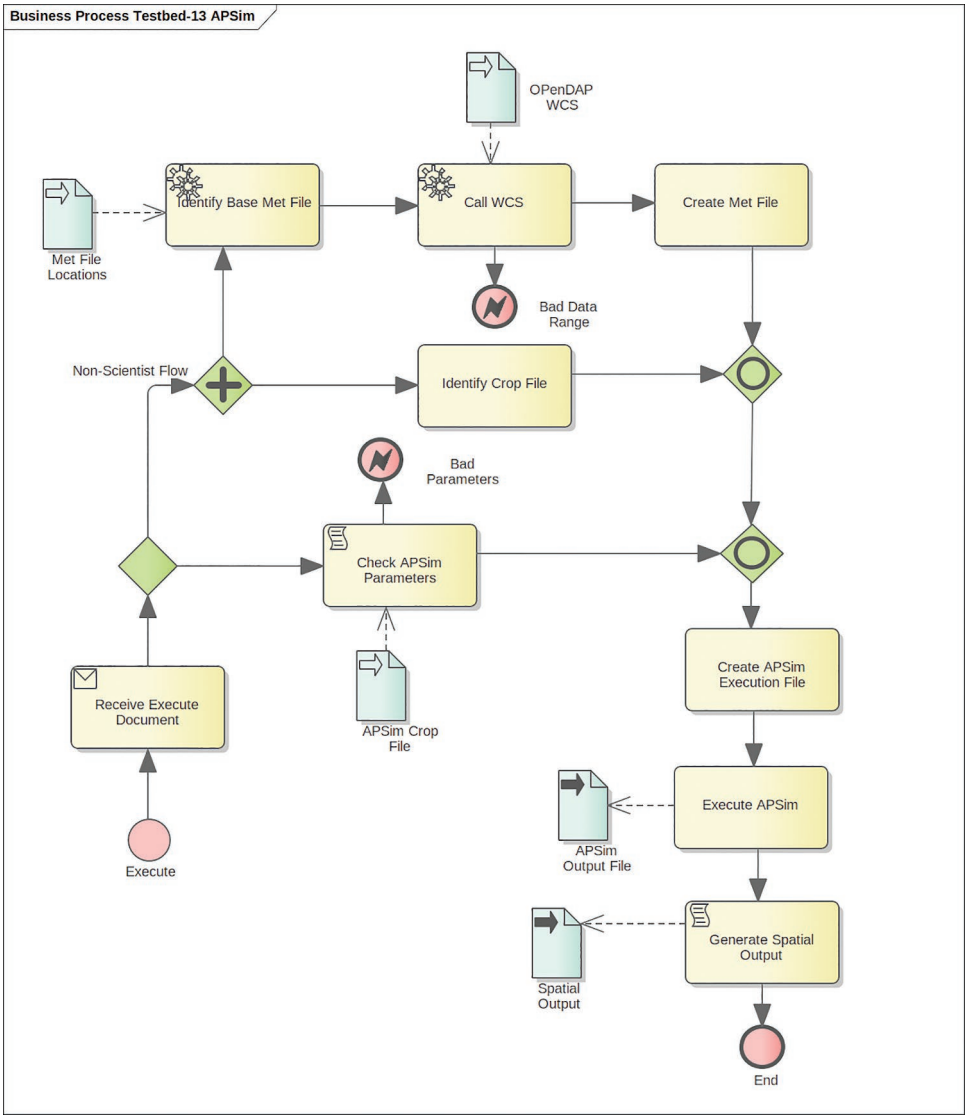

Figure 5.2 A BPMN diagram describing the workflow for the APSIM WPS
Source: OGC.

Note, in Figure 5.1 the FME environment plays the role of the BPMN workflow client and engine environment seen in Figure 5.2. Both make use of existing OGC standards to enable encapsulating within a WPS call of a combination of macro services such as APSIM. Within the GRASP project (Leibovici *et al.*, 2017a, 2017b) a BPMN workflow client was also designed to create the scientific modelling, combining lower-level transformations and computations themselves encapsulated in a WPS (Figure 5.3). In this type of diagram both data and processing are seen as OGC Web services. This enabled the combination of specific sub-models closer to designing bespoke agricultural scientific modelling as well as allowing the use of published services of established models.

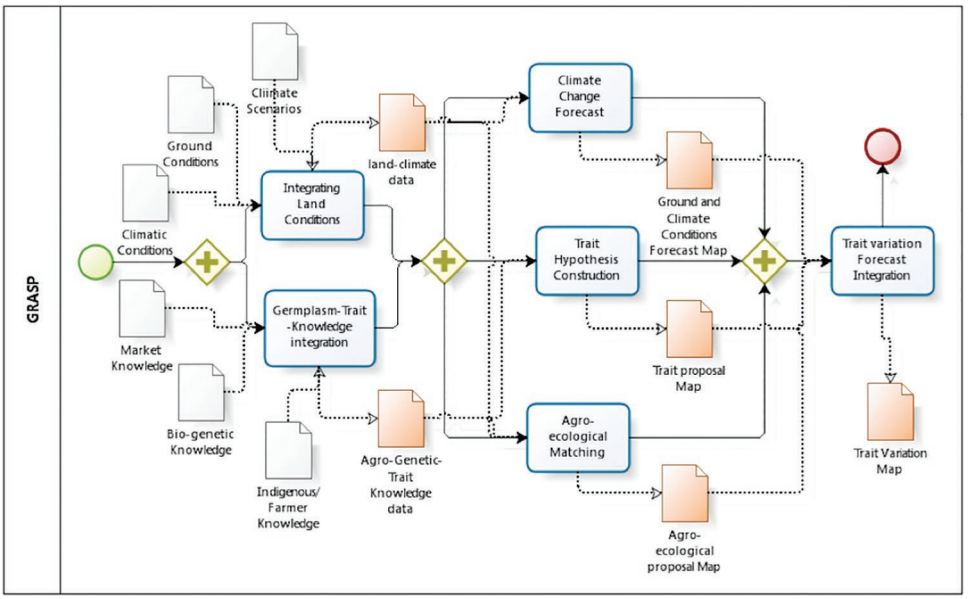

Figure 5.3 Workflow design of a generic model to be used in food security and sustainability: the Genetic Agro-ecological Sustainability Proposal model (BPMN diagram)

Source: Leibovici *et al.* (2017a, 2017b).

Changing granularity, each submodel can be represented as a BPMN workflow diagram combining even lower-level geospatial computational data transformations encapsulated again in WPS making use of appropriate APIs. In order to facilitate the combination of such lower-level processing facilities, metadata derived from using these APIs brings identification and quality within a standardization framework (see below under 'Future research for standard adoption').

The use and adoption of WPS and BPMN standards in scientific modelling are still in their infancy as the WPS 'libraries' are not yet well-developed towards providing generic processing available for re-use. In terms of processing environment, the complexity (e.g., in crop modelling) has certainly been an important slowing factor as it makes the development of new systems much dependent on existing ones, therefore re-using previous computing environment, i.e., often as monoblocs of written code programs somewhat like APSIM. As multidisciplinary modelling in the environmental domain develops and more and more interoperable geospatial standards become available, communities dealing with complexity such as in crop modelling making use of climate forecasting data and weather information (Jones *et al.*, 2019), genetic data and geo-phenotypic information (Leibovici *et al.*, 2017a, 2017b), smart farming (Colezea *et al.*, 2018)), farming practices and agro-ecological data (Bordogna *et al.*, 2016; McMenzie *et al.*, 2019), and others, all stand to benefit.

The complexity in such geoprocessing modelling justifies the focus on the role of workflow environments such as using the BPMN standard in the development of geoprocessing facilities within the virtual research e-infrastructure such as proposed by the European Open Science Cloud (EOSC), the Open Science Framework (OSF) and services infrastructures such as the GEOSS (Group on Earth Observation System of Systems) (Barker *et al.*, 2019; Enescu *et al.*, 2019).

Connection with national standards

The presence and use of cloud computing resulted and generated an incredible amount of structured and unstructured geographical data, which now represents a benefit or challenge depending on how it is harnessed. Such data call for the use and re-use of standards to maximize its potential for innovation as evidence for decision making at the national, regional and continental levels. Understanding the evolution of standards is then critical, particularly so for the productive sector of agriculture.

ISO/TC 211 publishes a family of international standards that promote an understanding and usage of geographic information to increase the availability, access, integration and sharing of geographic information. The interoperability of geospatially enabled computer systems allows a unified approach to address global ecological and humanitarian problems through geospatial e-infrastructures by then contributing to sustainable development and delivering country commitments to the Sustainable Development Goals (SDGs) identified by the United Nations.

The coordinated development implied here requires an understanding of organizations that use geopolitical decision-making at national, regional and continental levels. This helps support the unhindered flow of digital spatial information that relies on open systems for data delivery within rapidly evolving sectors. Once the organizational framework is in place, a cooperative research and development paradigm allows to evolve the ways of sharing sectoral information promoting administrative and technical support within technologically aligned paradigms. This supports the transfer of the benefits accruing to the private sector for the facilitation of adoption and attendant private sector investment. However, realizing this is not possible if governments do not lead the consumption of GIS solutions. Demonstrations of innovation uptake are at their best when they are led by governments, particularly in the areas of real-time communication deriving from data infrastructure. This then promotes the creation of market demand that the private sector only fortifies and takes the lead when the commercial benefits are established using government demand of services and associated standards.

It is also important to acknowledge that there is a need for expanding capacity development (especially in lower-income developing countries). This will help in integrating OGC/ISO standards with regional and national standards. For example, the INSPIRE directive identifies a number of themes that form a framework for Spatial Information across Europe. One of those themes is 'agricultural and aquaculture facilities'. To enable member nations to conform to the INSPIRE directive, the European Commission led initiatives to develop Technical Guidelines for Data Specifications based on each of the themes (the INSPIRE Data Specification on Agricultural and Aquaculture Facilities – Technical Guidelines are available at https://inspire.ec.europa.eu/Themes/137/2892).

Citizen science: a new paradigm

Over the last few years, citizen science has advanced across many disciplines. In the geospatial domain, the rise of volunteered geographic information (VGI), citizen science (CS) and crowdsourcing (including social media), either providing information on existing geospatial data or being part of a data collection from using a geolocated device, has seen rapid rise in interest. CS is the concept of opening part of scientific work to the general public. There exist a number of terms that are used as synonyms or subcategories like crowd science, crowdsourced science, or in relation to spatial information there is VGI, citizens sense or citizens as

sensors. Some of the most widely used applications in this domain include OpenStreetMap (https://www.openstreetmap.org) and Geo Wiki (https://www.geo-wiki.org/). Generally, CS is closely linked to the concept of Responsible Research and Innovation (RRI), but it has a number of similarities with other initiatives or Technik's like Living Labs, Science Shops, Hackathons, Datathons and IdeaThons.

To better understand the full potential of CS-captured data for agriculture, it is necessary to compare different techniques and possibilities of CS and to analyse how CS could be easily harmonized and integrated with other spatial information. The Farm Oriented Open Data in Europe (FOODIE) is a recent open data initiative based in Europe with ambitions to help agriculture across the globe. It aims to increase efficiency and open new opportunities for all involved in planning, growing and delivering food to the marketplace. The FOODIE project will enable farmers to provide their own data by easy-to-use crowdsourcing tools and applications that encapsulate the complexity of the underlying services technology. Mobile or Web applications offer Web-forms collecting specific information from farms with a wizard application allowing a farmer to configure it after a few 'clicks'. Then, the sensor system within a farm can send periodic observations to the FOODIE platform. Challenges and lessons learned from projects like FOODIE further understanding of why CS projects use existing standards or do not, and whether the need to continuously evolve creates barriers for the adoption of standards.

The discussion paper 'StandardizedStandardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data' (OGC 16–129 2016) provides some answers, as well as proposes a flexible framework: SWE4CS. It makes use of OGC O&M and SWE Common standards suite. The paper also makes use of RDF and SKOS to leverage the Linked Data capabilities of the Semantic Web. The European project COBWEB was one of the first projects to apply these standards for CS as well as focusing on integrating the data quality issues (Higgins *et al*., 2016; Leibovici *et al*., 2017a, 2017b). CS is a key enabler for widening participation (especially youth in low-income countries) in agriculture and improving the livelihoods of smallholder farmers. For example, in the European project LandSense, contributing data from farmers are combined with Copernicus remote sensing Sentinel-2 data as a service delivered on the farmer's mobile phone to inform them about geolocated crop health (CropSupport app), with the fusion of standardized information that enables this to happen.

The FOODIE consortium has found a consensus on the content and structure of the FOODIE Data Model. The FOODIE model is a complex model including Core model, Transport Data Model, FOODIE sensor observations profile and FOODIE VGI profile. Within FOODIE data modelling activities, a VGI profile will be developed, enabling farmers and other volunteers to add agri-based content on the top of the base map provided by the FOODIE platform. An example of such an approach is currently used in the United States for the drought monitoring program CLIMAS. The primary assumption is that a farmer or volunteer is not an expert, but has skills for monitoring. For that reason, registration of the volunteers on the FOODIE platform is useful, since reporting options will be then limited by the VGI profile that suits the skills of a volunteer. In that regard, the VGI profile for meteorological monitoring is foreseen as the initial profile.

VGI has often been used for reporting of negative features. Examples in the agriculture domain may be found in the reporting of drought, mildew, wrong application of a pesticide beyond the field, and so on. The FOODIE VGI profile was developed in mid-2016 within this consensus and integrated into the Sensor Observation profile as an additional extension (Figure 5.4), taking advantage of the tested structure and respecting the conception of universal observation with result value of any data type (Řezník *et al*., 2016).

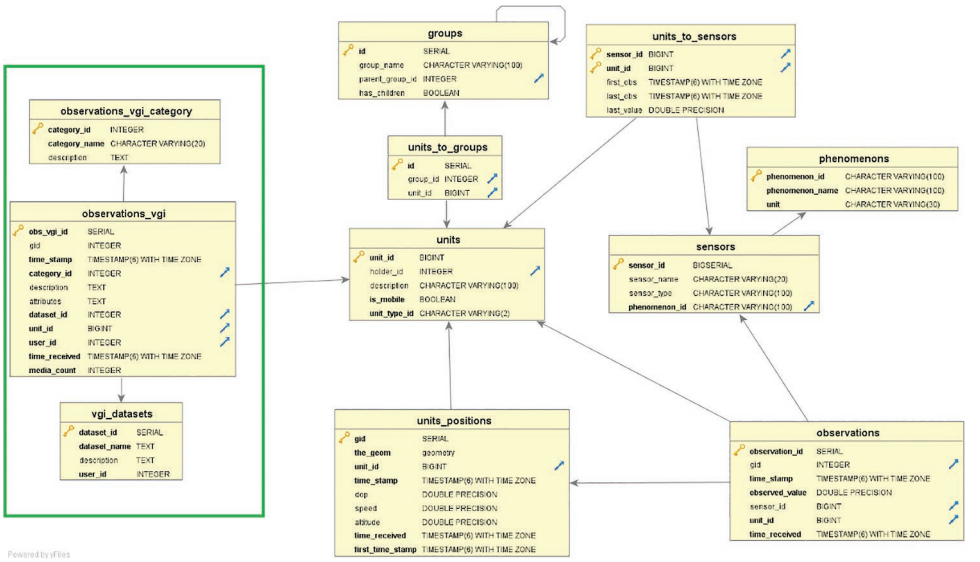

Figure 5.4 Snapshot of the FOODIE VGI profile
Source: Řezník *et al.* (2016).

There are many types of sensor data available on the Web with a variety of formats, encoding and standards. Wide-ranging services and solutions for publishing sensor data from different types of sensors are usually very complicated for non-expert users. The SDI4 Apps address this problem by designing tools and processes for effortless and straightforward utilization of existing sensor services available on the Web. This work has been done in the scope of the Open Sensors Network pilot as part of the SDI4 Apps project (Kepka and Ježek, 2012). The aim of Open Sensors Network is to create an environment where different groups of volunteers (e.g., farmers) can integrate low-cost sensors (meteorological data, air quality, and so on) into local and regional Web sensor networks. Re-using existing sensors operated by third-party providers is currently problematic. The problems range from initial search of services through metadata description of sensors to filtering of found sensor candidates. Therefore, a new module for VGI was designed and implemented. This VGI module contains new services for receiving and publishing data over the Web. The SensLog data model was extended with new tables with emphasis on variability of VGI. A VGI observation is characterized by several mandatory attributes and can be enriched with additional attributes (see Figure 5.5).

Crowdsourced, CS and VGI data, perhaps more than other data due to their recent integration into other SDIs, are likely to expose evolving solutions for data specification. Before convergence of these solutions is established, the cross-transformation from one data model to another would ensure as much working interoperability as possible while promoting its adoption. As this was already recognized in Charvat *et al.* (2014), integration of this type of data with existing SDIs (INSPIRE, GEOSS) can bring benefit to the development of both data models. On the other hand, it required action on both sides, i.e., by official SDIs to introduce lighter API-based interfaces and by CS to put effort into standardization to support easier integration.

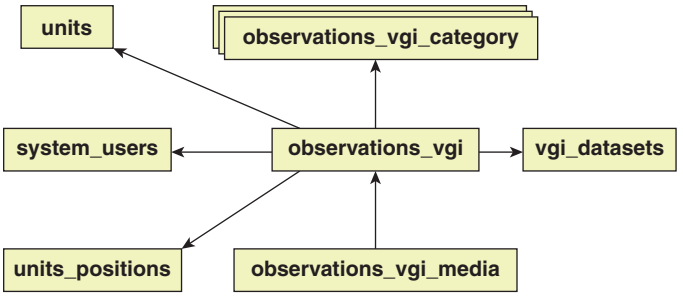

Figure 5.5 Revised FOODIE VGI profile of SDI4apps based on principles of Citizen Sensors
Source: Kepka and Ježek (2012).

Future research for standard adoption

Standards for automatic re-use of data by humans and machines is critical for efficient use of advancement in data science, such as with machine learning and artificial intelligence. Standards do not only propel the automatic re-use of data with interoperability for these emerging data science applications, but will also provide much better provenance information on how these systems are dealing with personal data (GDPR) and use licensed input data (CC-by, proprietary, and so on). The domain-specific communities in the Research Data Alliance are working on specification of domain standards.

Artificial intelligence, and more specifically machine learning, are current areas of interest for the geospatial community. Machine learning models are being developed and applied to geospatial data such as remote sensing imagery to improve feature extraction in support of land cover classification. An example of a project that examined the application of machine learning in geospatial solutions is the OGC Testbed-15 project, which used environmental data from Canada to explore scenarios such as:

- Forest change prediction model (Petawawa Super Site Research)
- Forest supply management decision-maker machine learning models (New Brunswick)
- River differentiation machine learning models (Quebec Lake)
- Linked data harvest models (Richelieu River Hydro)
- Arctic Web services discovery machine learning models

The outputs of the OGC Testbed-15 project are expected to inform future standardization activities. Given the importance of environmental management to agricultural operations, future research should look to apply emerging standards and best practices that integrate machine learning into geospatial solutions. These machine learning applications are often integrated within domain-specific models, and the scientific workflow representing the whole model can be seen as the metadata of the entire geoprocessing, playing a role in the Open Science approach. Open Science relevant to the various scientific communities involved in CODATA (the Committee on Data of the International Science Council or ISC), which exists to promote global collaboration to advance Open Science and to improve the availability and usability of data for all areas of research.

The virtual research environment and initiatives of open science, such as the European Open Science Cloud (EOSC), OSF (Open Science Framework) as well as the GEOSS (Group on Earth Observation System of Systems) in which standards and FAIR principles

are paramount, have educational content as well as propose a scientific dissemination and reproducible framework where the workflow is encoded (e.g., using the BPMN standard) and gives a visual semantic as well as executable capacity (via a workflow engine). Within such systems, the workflow itself plays a role as metadata that will need to be standardized, also containing the metadata of each part of the workflow (data or processing). These metadata are driving the analytics (e.g., derived from machine learning) that can be associated with the scientific workflow, extending it towards decision-making, including the attached uncertainty assessment. Eventually, the metadata composition in the workflow contains the mechanisms to run the workflow itself as well as these analytics (Rosser *et al.*, 2018).

The recent development of the OGC API standards is also modernizing the way through which geospatial data is accessed over the Web. An API is an interface that when backed by implementing code satisfies an interoperability need, as documented in a Requirements Specification. Web APIs use an architectural style that is founded on the technologies of the Web (W3C), which has led OGC to develop a series of new specifications, the OGC APIs. The approach taken in the OGC APIs is consistent with the W3C Data on the Web Best Practices. For example, the OGC API – Features standard, approved in October 2019, allows for the dissemination, query and retrieval of geospatial feature data. Documentation about OGC APIs is provided through OpenAPI documents (or its commercial equivalent, Swagger). In the foreseeable future, agricultural solutions may offer implementations of OGC API – Features that disseminate data encoded according to SoilML. A data infrastructure for farming, for instance, could also deploy implementations of the OGC SensorThings API standard to serve observations collected by sensors on the farm. The OGC SensorThings API standard allows for sensors in the Internet of Things to be accessed in order to retrieve observations. For requesting entities, the OGC SensorThings API follows the OData specification of the Organization for the Advancement of Structured Information Standards (OASIS). Future research should explore how such an API-driven architecture could improve the effectiveness of farming operations, agriculture modelling and decision-making.

Conclusion

Currently the standards community in the geospatial domain is experiencing a strong incentive to capitalize on the long history of standards through the advancement of open science and more specifically the adoption of FAIR as the guiding principles for better reuse of (research) data by humans and machines. It is critical that the geospatial community recognizes its existing role as an area where standards are re-used and engages with other scientific communities to re-use the existing standards in, for example, the health sector or in the logistics sector.

Research data scientists in CODATA, GO FAIR, RDA and others are proposing a globally generic standard on how to deal with FAIR Digital Objects in order to establish FAIR compliant e-infrastructures (e.g., Wittenburg *et al.*, 2019). Initiatives such as EOSC and CODATA should be encouraged because they go in the direction of increasing the adoption of the long list of existing geospatial standards and into design of future standards that are able to answer new challenges.

References

Barker, M., Olabarriaga, S.D., Wilkins-Diehr, N., Gesing, S., Katz, D.S., Shahand, S., Henwood, S., Glatard, T., Jeffery, K., Corrie, B. and Treloar, A. (2019) "The Global Impact of Science Gateways, Virtual Research Environments and Virtual Laboratories" *Future Generation Computer Systems* 95 pp.240–248 DOI:10.1016/j.future.2018.12.026.

Batjes, N.H. and van den Bosch, H. (2018) "How Well is the WDC-Soils Serving its User Groups?" *SciDataCon-IDW 2018* Gaborone, Botswana DOI:10.17027/6F38-HN42.

Bordogna, G., Kliment, T., Frigerio, L., Brivio, P. A., Crema, A., Stroppiana, D., Boschetti, M. and Sterlacchini, S. (2016) "A Spatial Data Infrastructure Integrating Multisource Heterogeneous Geospatial Data and Time Series: A Study Case in Agriculture" *ISPRS International Journal of Geo-Information* 5 (5) 73 DOI:10.3390/ijgi5050073.

Charvat, K., Esbri, M.A., Mayer, W., Campos, A., Palma, R. and Krivanek, Z. (2014) "FOODIE – Open Data for Agriculture" *Proceedings of the IST-Africa Conference* Pointe aux Piments, Mauritius, pp.1–9 DOI: 10.1109/ISTAFRICA.2014.6880647.

Coetzee, S., Ivánová, I., Mitasova, H. and Brovelli, M. A. (2020) "Open Geospatial Software and Data: A Review of the Current State and A Perspective into the Future" *ISPRS International Journal of Geo-Information* 9 (2) 90 DOI:10.3390/ijgi9020090.

Colezea, M., Musat, G., Pop, F., Negru, C., Dumitrascu, A. and Mocanu, M. (2018) "CLUeFARM: Integrated Web-Service Platform for Smart Farms" *Computers and Electronics in Agriculture* 154 pp. 134–154 DOI:10.1016/j.compag.2018.08.015.

Enescu, I.I., Plattner, G.K., Bont, L., Fraefel, M., Meile, R., Kramer, T., Espona-Pernas, L., Haas-Artho, D., Hägeli, M. and Steffen, K. (2019) "Open Science, Knowledge Sharing and Reproducibility as Drivers for the Adoption of FOSS4G in Environmental Research" *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4/W14 pp.107–110 DOI:10.5194/isprs-archives-XLII-4-W14–107–2019.

Guan, W.W., Bol, P.K., Lewis, B.G., Bertrand, M., Berman, M.L. and Blossom, J.C. (2012) "WorldMap – A Geospatial Framework for Collaborative Research" *Annals of GIS* 18 pp.121–134 DOI:10.1080/19475683.2012.668559.

Helbig, C., Dransch, D., Böttinger, M., Devey, C., Haas, A., Hlawitschka, M., Kuenzer, C., Rink, K., Schäfer-Neth, C., Scheuermann, G., Kwasnitschka, T. and Unger, A. (2017) "Challenges and Strategies for the Visual Exploration of Complex Environmental Data" *International Journal of Digital Earth* 10 pp.1070–1076 DOI:10.1080/17538947.2017.1327618.

Higgins, C.I., Williams, J., Leibovici, D.G., Simonis, I., Davis, M.J., Muldoon, C. and O´Grady, M. (2016) "Citizen OBservatory WEB (COBWEB): A Generic Infrastructure Platform to Facilitate the Collection of Citizen Science data for Environmental Monitoring" *IJSDIR* 11 pp.20–48.

Hoffmann, C., Schulz, S., Eberhardt, E., Grosse, M., Stein, S., Specka, X., Svoboda, N. and Heinrich, U. (2019) "Data Standards for Soil- and Agricultural Research" *BonaRes Data Centre* DOI:10.20387/BonaRes-ARM4–66M2.

ISO: ISO 19136:2007 "Geographic Information – Geography Markup Language (GML)" Available at: https://www.iso.org/standard/32554.html.

ISO: ISO 19142:2010 "Geographic Information – Web Feature Service" Available at: https://www.iso.org/standard/42136.html.

ISO 19156:2011 "Geographic Information – Observations and Measurements" Available at: https://www.iso.org/standard/32574.html.

ISO: ISO 28258:2013 "Soil Quality – Digital Exchange of Soil-related Data" Available at: https://www.iso.org/standard/44595.html.

Jones, A., Edwards, B., Reusch, K. and Fraccaro, P (2019) "Geospatial Big Data Analytics for Agriculture Using IBM-PAIRS" *Geophysical Research Abstracts* 21 Munich: European Geosciences Union.

Kempen, B., Yigini, Y., Viatkin, K., de Jesus, J.M., de Sousa, L.M., van den Bosch, R. and Vargas, R. (2019) "The Global Soil Information System (GloSIS) – Concept and Design" *Geophysical Research Abstracts* 21 Munich: European Geosciences Union.

Kepka, M. and Ježek, J. (2012) "Server-side Solution for Sensor Data" In Mildorf, T. and Charvat, K. (Eds) *ICT for Agriculture, Rural Development and Environment Where We Are? Where Will We Go?* Prague: Czech Centre for Science and Society, pp.264–274.

Leibovici, D.G., Anand, S., Santos, R., Mayes, S., Ray, R., Al-Azri, M., Baten, A., King, G., Karunaratne, A., Azam-Ali, S. and Jackson, M. (2017a) "Geospatial Binding for Transdisciplinary Research in Crop Science: The GRASPgfs initiative" *Open Geospatial Data, Software and Standards* 2 (20) pp.1–11 DOI:10.1186/s40965-017-0034-3.

Leibovici, D.G., Williams, J., Rosser, J.F., Hodges, C., Chapman, C., Higgins, C. and Jackson, M. (2017b) "Earth Observation for Citizen Science Validation, or, Citizen Science for Earth Observation Validation? The Role of Quality Assurance of Volunteered Observations" *Data* 2 (4) 35 DOI:10.3390/data2040035.

Montanarella, L., Wilson, P., Cox, S., Mcbratney, A., Ahamed, S., McMillan, B., Jacquier, D. and Fortner, J. (2010) "Developing SoilML as a Global Standard for the Collation and Transfer of Soil Data and Information" *Geophysical Research Abstracts* 12 Munich: European Geosciences Union.

OASIS (2014) "OData Version 4.0 Part 1: Protocol" Available at: http://docs.oasis-open.org/odata/odata/v4.0/os/part1-protocol/odata-v4.0-os-part1-protocol.html.

OGC 16–129 (2016) "Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data" *OGC Discussion Paper* Available at: http://www.opengis.net/doc/DP/16-129.

OGC & W3C (2017a) "Semantic Sensor Network Ontology" Available at: https://www.w3.org/TR/vocab-ssn/.

OGC & W3C (2017b) "Spatial Data on the Web Best Practices" Available at: https://www.w3.org/TR/sdw-bp/.

OGC & W3C (2022) "Time Ontology in OWL" Available at: https://www.w3.org/TR/owl-time/.

Pourabdollah, A., Leibovici, D.G., Simms, D.M., Tempel, P., Hallett, S.H. and Jackson, M.J. (2012) "Towards a Standard for Soil and Terrain Data Exchange: SoTerML" *Computers & Geosciences* 45 pp.270–283 DOI: 10.1016/j.cageo.2011.11.026.

Řezník, T., Lukas, V., Charvat, K., Charvat Jr, K., Horáková, Š. and Kepka, M. (2016) "Foodie Data Models for Precision Agriculture" *13th International Conference on Precision Agriculture* St. Louis, MO, 31th July–4th August, 2016.

Ritchie, A. (2016) "OGC Soil Data Interoperability Experiment" *Open Geospatial Consortium*, No. OGC 16-088r1" Available at: https://portal.opengeospatial.org/files/?artifact_id=69896.

Rosser, J., Jackson, M.J. and Leibovici, D.G. (2018) "Full Meta Objects for Flexible Geoprocessing Workflows" *Transactions in GIS* 22 (5) pp.1221–1237 DOI:10.1111/tgis.12460.

Schumann, G. (2018) "OGC Testbed-13: NA001 Climate Data Accessibility for Adaptation Planning" No. OGC 17-022 *Open Geospatial Consortium* Available at: http://docs.opengeospatial.org/per/17-022.pdf.

W3C (2017) "Data on the Web Best Practices" Available at: https://www.w3.org/TR/dwbp.

Wilson, P., Simons, B. and Ritchie, A. (2013) "Opportunities for Information Model Driven Exchange and On-line Delivery of GlobalSoilMap Data and Related Products" *Proceedings of the 1st GlobalSoilMap Conference: Basis of the Global Spatial Soil Information System* DOI:10.1201/b16500-85.

Wittenburg, P., Strawn, G., Mons, B, Boninho, L. and Schultes, E. (2019) "Digital Objects as Drivers towards Convergence in Data Infrastructures" *EUDAT b2share records* DOI:10.23728/b2share.b605d85809ca45679b110719b6c6cb11.