EBONE

European Biodiversity Observation Network:

Design of a plan for an integrated biodiversity observing system in space and time

D7.1: Technical Specification

Ver 1.0

Document date: 2010-03-30

Document Ref.: EBONE-D7.1_1.0_TechnicalSpecification

Authors: J. Peterseil, Blankman, D., Gaigalas, G., Magagna, B., Schentz, H., Vanacker, S., Vanhercke, L., van der Werf, B.

EC-FPV Contract Ref: ENV-CT-2008-212322

< Blank Page >

D7.1: Technical specification

Content

1 Purpose of the document 7
2 Introduction 7
2.1 INSPIRE 8
2.2 GEOSS Data Sharing Principles 9
2.3 LifeWatch 10
2.3.1 Reference Model 11
2.4 Impact of existing initiatives for EBONE 13
3 Requirements for EBONE 13
4 Data specification 14
4.1 Data sources 14
4.2 Data levels 14
5 Technical specification 15
5.1 Architecture 15
5.2 Data flow 18
5.3 Data model 19
5.3.1 Definition of terms 19
5.3.2 OGC Observation & Measurement 22
5.3.3 SERONTO 26
5.3.4 Implementation 28
6 Implementation strategy and plan 29
6.1 Metadata Management Component 29
6.1.1 Importance of extensibility 30
6.1.2 Types of Metadata 30
6.1.3 Structural / Semantic 31
6.1.4 Structure of Metadata 31
6.1.5 Implementation 32
6.2 Source Data Management Component 32
6.2.1 MONBEL - Field computer application (INBO) 33
6.2.2 Field database CEMAGREF 35
6.2.3 Analysis result data 39
6.3 Data Integration Component 41
6.3.1 Web services 42
6.3.2 Data warehouse 46
6.3.3 Implementation of the common domain model 49
6.4 Data Presentation Component 50
6.4.1 Context 50
6.4.2 Application (EFDAC Map Viewer) 51

6.4.3 Software, interface ... 54
6.4.4 User & Functional – technical requirements ... 60
6.4.5 Implementation ... 66
6.5 Helpdesk ... 66
6.6 Cost aspects ... 66
6.7 Time plan ... 66

7 First implemented steps ... 67

8 Test cases existing data ... 68
8.1 NILS ... 68
8.1.1 Data structure ... 69
8.1.2 Data management ... 69
8.1.3 Data transformation ... 70
8.1.4 Data policy issues ... 70
8.2 SINUS ... 70
8.2.1 Data structure ... 72
8.2.2 Data management ... 72
8.2.3 Data transformation ... 72
8.2.4 Data policy issues ... 74
8.2.5 Data example ... 74
8.3 SISPARES - The Spanish Rural Landscape Monitoring System ... 75
8.3.1 Data structure ... 75
8.3.2 Data management ... 76
8.3.3 Data transformation ... 76
8.3.4 Data policy issues ... 76
8.4 British Country Side Survey / North Ireland Country Side Survey ... 76

9 Outlook ... 77

10 References ... 77

Version	Date	Author	Note
0.1 draft	03.03.2010	Peterseil	Compiling the contribution into the report structure
0.2 draft	22.03.2010	Magagna	Include EFDAC part and O&M part
0.3. draft	24.03.2010	Peterseil	Include contributions to LifeWatch, data integration
0.4 draft	25.03.2010	Vanhercke / Blankman	comments
0.5 draft	30.03.2010	Vanacker, Blankman, Schentz	Review
1.0	31.03.2010	Peterseil	Final check – Final version

Summary

Data management is a core function for monitoring and research activities. The integration of existing data sources as well as the structure of newly collected data is a major task. The aim of EBONE is to provide a INSPIRE and GEOSS compliant data management structure which allows the access and use of distributed data sources.

Many European initiatives aim in the integration of existing data and provide reference models, concepts and tools. For EBONE the appropriate tools were selected to enable a efficient and low cost data management based on existing tools and experiences.

The EBONE data management system consists of the following components:

a) Metadata Management Component
b) Source Data Management Component
c) Data Integration Component
d) Data Presentation Component

A Data Analysis Component will be needed to automate the calculation of biodiversity indicators for the EBONE. This component is not in the focus of the current report and will be developed with EBONE WP3.

This includes the following activities:

- Provide data entry tools for the field collection campaign
- Definition of a common domain model and application schema
- Set up web services for the existing data sources
- Implementation of a data warehouse for the integration of existing data sources

EBONE aimed to use open source and standardised tools to implement this system. The implementation will be tested in 2010 and an updated report will be provided by the end of the year.

1 Purpose of the document

The purpose of the document is to lay out the components for the EBONE (European Biodiversity Observation Network) data management framework and the strategy for their implementation. In this sense the current document is termed as technical specification.

The technical specification is considered as a living document which is going to evolve during the lifetime of the EBONE project. Its aim is to present the current status of the common view on the system architecture and components of the data management framework. According to the project evolvement the different parts of the document and the concepts need to be revised and updated.

The present status of the document reflects the current status of discussion. Work on the refinement of the system architecture will be done during and after the test phase within the runtime of the project.

The basic elements of the concept will be tested during 2010 and an updated implementation strategy will be produced by beginning of 2011.

2 Introduction

Data management is a key feature in a monitoring framework. One of the core tasks is to provide the data with sufficient metadata for analysis and calculation. The EBONE data management framework should build on existing tools and will adopt these according to the needs for a European Biodiversity Monitoring framework. The goal of the EBONE work package on data management is to provide a **data management system**, which allows storing standardised parameters and methods for the European Biodiversity monitoring network. This includes the following tasks: a) to determine essential **operational core services**, b) to **determine relevant data flows and data** according to INSPIRE and GEO data sharing principles, c) to **provide a database for the collected field test data**, and d) to **design data architecture and technical tools** for needed services.

The development and the analysis of databases are essential for the project. The following prerequisites are needed for databases and linking data:

- Structured data storage (in central or dispersed data bases)
- Standards used for the description of methods and parameters
- Data available free of charge or reasonably priced
- IT-knowledge available in institutions or their environment

On the European scale initiatives try to solve the problem of data interoperability in the biodiversity domain (e.g. Alter-Net, LifeWatch). Within the current project this experience was used to identify the scope of the needed components for a biodiversity information management framework dealing with habitat and species monitoring.

Based on the analysis of the structures of existing databases and data sources and the user needs for data integration a concept of the needed components for the information management framework was formulated. In the current report these components are described and a strategy for the implementation for a test case is proposed.

2.1 INSPIRE

The infrastructure for Spatial information in the European Community (INSPIRE) is an initiative of the European Commission to establish a European spatial information infrastructure as a basis for services providing access to spatial information held by public authorities. The INSPIRE Implementation rules detail the INSPIRE legislative requirements of data and metadata so that these can be implemented consistently across Europe. EBONE shall aim to conform with the INSPIRE Implementation Rules so that EBONE can provide and access all INSPIRE conformant data and services.

INSPIRE is a Directive proposed by the European Commission in July 2004 setting the legal framework for the establishment and operation of an Infrastructure for Spatial Information in Europe. The purpose of such infrastructure is to support the formulation, implementation, monitoring activities and evaluation of Community policies linked with the environment at all levels, European, national and local, and to provide public information.

INSPIRE should be based on the infrastructures for spatial information that are created by the Member States. The components of those infrastructures include: metadata, spatial data themes (as described in Annexes I, II, III of the Directive), spatial data services; network services and technologies; agreements on sharing, access and use; coordination and monitoring mechanisms, processes and procedures. The guiding principles of INSPIRE are:

- that the infrastructures for spatial information in the Member States should be designed to ensure that spatial data are stored, made available and maintained at the most appropriate level;
- that it is possible to combine spatial data from different sources across the Community in a consistent way and share them between several users and applications;
- that it is possible for spatial data collected at one level of public authority to be shared between all the different levels of public authorities;
- that spatial data are made available under conditions that do not unduly restrict their extensive use;
- that it is easy to discover available spatial data, to evaluate their fitness for purpose and to know the conditions applicable to their use.

The text of the INSPIRE Directive is available from the INSPIRE web site (http://www.ec-gis.org/inspire).

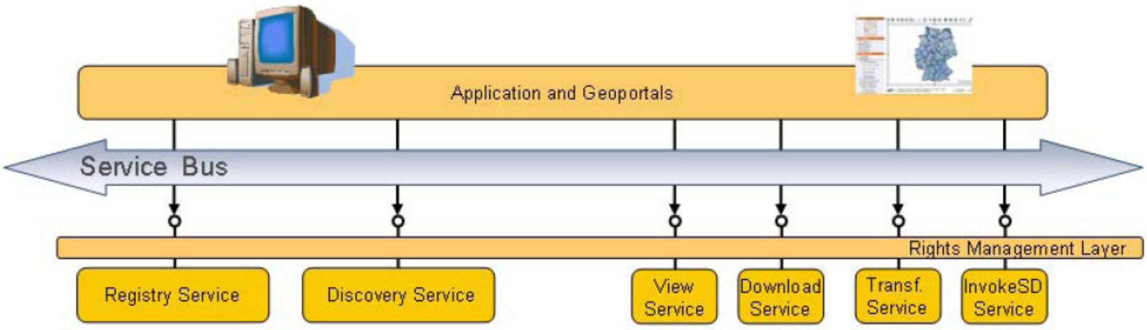


Fig. 1 INSPIRE Service Types (Network Service Drafting Team, 2008)

The goal of the **discovery service** is to support discovery of data, evaluation and use of spatial data and services through their metadata properties (Network Service Drafting Team, 2008). Metadata is the information and documentation, which makes these resources understandable and shareable for users over time.

The **view services** allow to display, navigate, zoom in/out, pan or overlay viewable spatial data sets and to display legend information and any relevant content of metadata.

A **download service** support the download of a complete dataset or datasets, or a part of a dataset or datasets, and where practicable, provides direct access to complete datasets or parts of datasets. Transformations between the application schema of the spatial dataset and the harmonised schema may be done on-the-fly or by **transformation services**.

The **invoke spatial data service** allows the definition of both the data inputs and data outputs expected by the spatial service and define a workflow or service chain combining multiple services. It also allows the definition of a web service interface managing and accessing (executing) workflows or service chains. For spatial data services available on the Internet, the "Invoke Spatial Data Service" service will enable a user or client application to run them without requiring the availability of a GIS. This requires that a client application can discover the service, bind to it and invoke it. The orchestration/combination of Spatial Data Service with other services will require to precisely defining the interactions between the services. Therefore, the interaction between the (spatial) services to be invoked is defined as a workflow or composite service in a standard notation (e.g. XML-based).

The general aim of the INSPIRE architecture is to store and manage data as near as possible at the organisation where the data are managed and provide view and access to this dataset via services.

2.2 GEOSS Data Sharing Principles

EBONE should also include the GEOSS data sharing principles as a core for the information management. The GEOSS 10-Year implementation explicitly acknowledges the importance of data sharing in achieving the GEOSS vision and anticipated societal benefits. The plan, endorsed by nearly 60 governments and the European Commission at the 2005 Third Earth Observation Summit in Brussels, highlights the following GEOSS Data Sharing Principles:

- There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;
- All shared data, metadata and products will be made available with a minimum time delay at minimum cost;
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

2.3 LifeWatch

LifeWatch[1] will construct and bring into operation the facilities, hardware, software and governance structures for all aspects of biodiversity research. It will consist of: facilities for data generation and processing; a network of observatories; facilities for data integration and interoperability; virtual laboratories offering a range of analytical and modelling tools; and a Service Centre providing special services for scientific and policy users, including training and research opportunities for young scientists. The infrastructure is supported by all major European biodiversity research networks.

The LifeWatch ICT Infrastructure (or infrastructure for short) shall be a distributed system of nodes that provide access to and processing of biodiversity data from a variety of sources through common open interfaces for several decades.

In a technical construction plan the strategy how this shall be achieved is specified and in the reference model the reference architecture and it's components are described. A Reference Model is an abstract framework for understanding the significant relations among the entities of the subject of concern. Its purpose is to provide a common conceptual framework defining the key characteristics that can be used consistently across different implementations. Which means the implementation details and technology can still be freely chosen (Hardisty, 2010).

LifeWatch tries to tackle the following issues:

- A large gap between the ICT-based research practices in common/widespread use across the biodiversity research community today, and the vision of LifeWatch as a future interoperable 'e-Science' infrastructure offering collaborative facilities to groups of scientists.
  Progress towards solving this challenge requires engagement with open-minded scientists willing to engage, explore and progress.
- The pace of ICT innovation is rapid, making it hard both to specify a stable platform that meets the needs of scientists and to "home in" on solutions with potential to achieve the LifeWatch vision.
  Progress towards solving this challenge is in part accomplished by the LifeWatch Reference Model (LifeWatch-RM). LifeWatch-RM is a mechanism for making technology independent design decisions now, for the basis of the preparatory work, which can be instantiated later-on using specific technology approaches.

[1] http://www.lifewatch.eu/

These approaches can be supported both by an appropriate technical governance model, based on the meritocratic approach of the Apache Software Foundation, and the strategy set out in the present document. This is called "European strategies for local implementation".

2.3.1 Reference Model

LifeWatch will construct an environment where many different initiatives interoperate to provide rich mechanisms for biodiversity knowledge exploration, analysis, and discovery. Major capabilities are (taken from Hernandez-Ernst et al., 2010):

- An interoperable platform to support the exchange and use of services for discovery, data access and data analysis
- The provision of virtual research environments, which allow the dedicated use of the services
- Provision of the appropriate use of resources through security mechanisms and IPR and attribution policies

These capabilities can be provided, considering the following requirements:

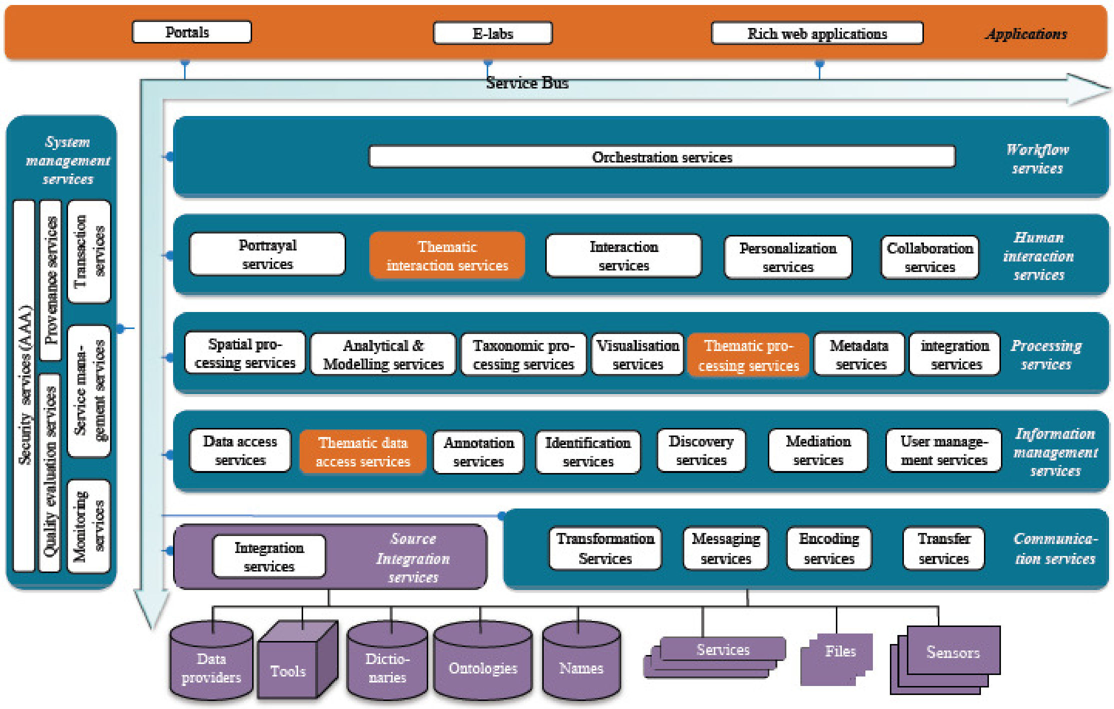
- Interoperability
    - Standardised data formats, which are independent of specific data models and representations
    - Standardised protocols or interface specifications, which describe services independent from a particular implementation
    - Global identifiers, (based on global naming conventions and terms) which persist independent from a particular technology or resolution process
    - Semantic mediation
- Virtual research environments
    - Combination of services through workflows
    - Provision of distributed computing e.g. through GRID environments
    - Provision of provenance information
- Security, privacy and attribution aspects
    - Use of control mechanism (e.g. from GRID, AAA-Services)
    - Ensure appropriate citation mechanisms (e.g. through provenance and GUID's)

The LifeWatch Reference Model provides a blueprint for organising these capabilities into a consistent framework, giving guidance to all stakeholders and developers, and it provides a set of rules developers should adhere to in order to conform to the requirements of the LifeWatch ICT infrastructure. The very basic assumption is that the LifeWatch ICT infrastructure will be constructed according to the principles of a semantically enhanced Service Oriented Architecture (SOA) (Hernandez-Ernst et al., 2010).

A Reference Model is an abstract framework for understanding the significant relations among the entities of the subject of concern. Its purpose is to provide a common conceptual framework defining the key characteristics that can be used consistently across different implementations. A Reference Model should be distinguished from a Reference Architecture that serves as an abstract model from which concrete architectures can be derived. The Reference Architecture focuses on the components of a system and their relationships, introducing concepts and architectural elements as needed in order to fulfil core requirements of the system to be constructed, but avoiding reliance on specific technologies.

The LifeWatch Reference Model may be considered as an intermediate between Reference Model and Reference Architecture as discussed above. It provides a common conceptual framework as well as it defines a number of components and architectural concepts as a basis for the future LifeWatch Architecture. It is neither a blueprint nor does it define a technological mapping, but identifies some key aspects and components that should be present in the final implementation of the LifeWatch System.

General overview of scope and structure of the LifeWatch architecture, all its components:



The LifeWatch Reference Model adapts and extends the ORCHESTRA Reference Model. It is based on many standards such as those advanced by OASIS (Organization for the Advancement of Structured Information Standards), OGC (Open Geospatial Consortium), by the organisation for Biodiversity Information Standards (TDWG), as well as those emerging in the biodiversity research community.

It is planned to complement the LifeWatch Reference Model with a "How-to" guide that will provide examples of how to interpret and apply the reference model as guidance for developers.

2.4 Impact of existing initiatives for EBONE

As LifeWatch is not yet in the construction phase, EBONE can not yet rely on an existing LifeWatch infrastructure but rather can do anything to prepare for such an infrastructure to come in future years.

This means that EBONE infrastructure follows LifeWatch Reference Model:

- Interoperability
    - EBONE will make use of OGC Standards for data formats and protocols
    - EBONE will provide a sort of basic semantic mediation by the use of ETL (extract – **translate –** load) for the data warehouse with the help of which data analysis will be done.
      Whenever LifeWatch will provide specific software it will be easy to do necessary shifts as the principles of the reference model are already followed.
- Virtual research environments
    - A virtual research environment is provided by the use of a data warehouse . Better virtual research environments are highly welcome whenever they are provided.
- Security, privacy and attribution aspects
    - The same as with the virtual research environment: privacy and IPRs are so far treated at a very low level and will be changed as soon as appropriate software will be available.

3 Requirements for EBONE

For the EBONE data management framework the following use cases can be defined. A more detailed report on the requirement collection will be provided with the updated version of the technical implementation:

- Find Biodiversity related data sets (habitats and species) within the existing data storages and the newly recorded data. – Presented as a standardized metadata discovery service → *Metadata Management Component*
- Management of field observation data (Field recording) : Use a mobile data management system for recording data, measured and observed in the field according to the EBONE standards .→ *Data Source Component (field data management)*
- Management of field observation maps (ArcGIS): Use a mobile data managements system for recording geographical data in maps according to EBONE standards→ *Data Source Component (field data management)*
- Access to existing biodiversity relevant data sources which comply to the EBONE field recording → *Data Source Component (existing distributed data sources)*
- Link local distributed data sources with biodiversity relevant data which comply to the EBONE field recording to the central data repository (data ware house) → *Data Integration Component*
- Download biodiversity related data (habitats and species) → *Data Integration Component (with a standardized data download service)*

- Visualise biodiversity related data sets (habitat and species) → *Data Presentation Component (web (map) viewing service)*
- Extract harmonised data for scientific analysis from distributed data sources - analysis, right management, access policy → *Data Presentation Component*
- Calculate biodiversity indicators based on downloaded data or data integrated by the central data repository → *Data Analysis Component*
- Extract harmonised data for analysis from distributed data sources → *Data Analysis Component (indicator calculation)*

4 Data specification

The EBONE information management framework has to deal with habitat and species monitoring data, as well as data from earth observation dealing with habitat and species occurrence. In the following chapter the data sources within EBONE are described and explained.

4.1 Data sources

Different types of data sources can be identified for the EBONE network the data management has to deal with:

i. Field data
   - Data mapped according to the EBONE mapping procedure (GHC/species) on new sites. These data are full compliant with the EBONE data structure and raw data should be available in most of the cases. This data originate either from test mapping activities within the EBONE project or in a later stage from implementations from the EBONE habitat and species monitoring protocol on the national or regional level.
   - Data from existing monitoring schemas which are harmonised and transformed according to the EBONE transformation rules for GHC/species. These data are based on different data models which have only a certain level of compliancy to the EBONE data structure. Furthermore often raw data and their metadata can not be directly accessed but only aggregated values for different parameters for a defined analysis unit (e.g. landscape squares) are available.

ii. Earth observation data
   - Land cover classifications or other remote sensing products (e.g. phenology, fragmentation

The data management in EBONE has to be able to deal with different data characteristics and take into account aspects of data policy and data rights

4.2 Data levels

The data management system also has to address data on different levels, which can be distinguished as

- Raw field data on the level of the landscape square. These are the mapped data (e.g. GHC or other habitat classification according to the mapping protocol) together with their exact location and shape (spatial information).

- Aggregated data on the level of the landscape square. These are transformed (according to the GHC) and aggregated values, e.g. as sum of area (or proportion) of habitat categories or species per landscape which are the basis for further calculation. The exact spatial location of the landscape element within the landscape square is not provided. In some cases not even the exact location of the landscape square is provided but only the assignment to a Environmental strata or zone.
- Aggregated data on the level of the reporting unit (e.g. Environmental Strata and Zones). The Environmental stratification forms the basis for the calculation of the indicator values. Therefore this data level is based on aggregated figures of selected indicators based on the entry values of data level II for the Environmental Strata or Zone. Theoretically every other reporting unit is possible if the data meet the statistical requirements for the calculation of the indicator values for this reporting unit.

5 Technical specification

The following chapter describe the intended architecture and steps for the implementation. The components of the intended system are described and an outlook on the implementation in the test phase is given.

5.1 Architecture

The aim for EBONE is to establish a simple and INSPIRE compliant architecture. This should be based on existing applications and tools. Different tasks of the framework can be distinguished: a) data entry, b) data management, and c) data presentation. These reflect also the main use cases for EBONE.

The data analysis to calculate and derive biodiversity relevant indicators is another one which is positioned outside of the data management framework. It will be based on downloaded data. Only the resulting indicator values will be stored for further use and presentation.

Fig. 2 gives an overview of the elements of the data management framework. For each element the most appropriate tool must be chosen and integrated to a coherent data management framework. The starting points for the evaluation of the data management solutions are defined based on an evaluation of existing tools, applications and experiences within the EBONE consortium.

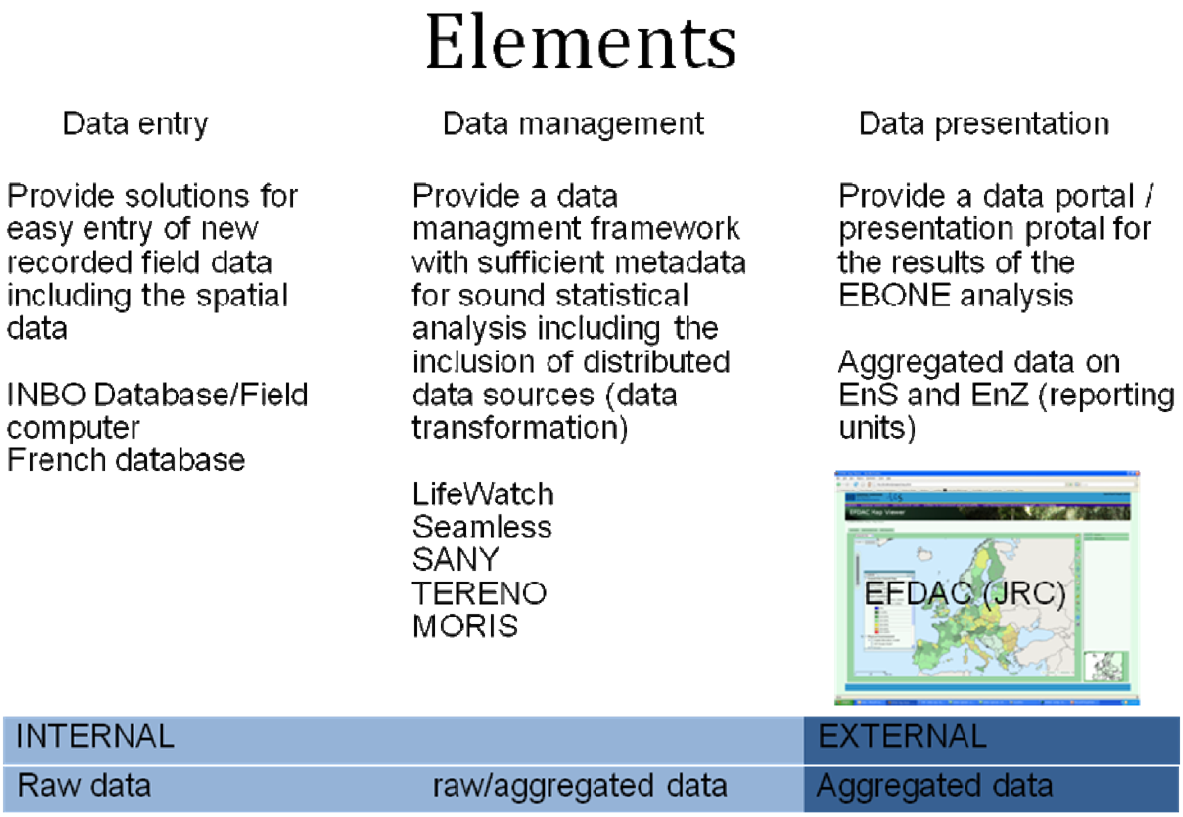


Fig. 2 Elements of the EBONE data management

Based on this use cases or elements of the EBONE data management the components of the system architecture can be defined. These are:

a) Source Data Management Component
    a. the field database (including the field computer application), for newly collected data
    b. existing data sources (e.g. ongoing monitoring activities)
b) Metadata Management Component (e.g. GeoNetwork or similar tool),
c) Data Integration Component (data warehouse, WFS),
d) Data Presentation Component (WMS/WFS Client).
e) Data Analysis Component (not included in the current document)

Fig. 3 gives an overview of the components of the EBONE system architecture. The *field database* contains the data collected according to the EBONE field manual (Bunce et al., 2010). The data model is structured according to the EBONE field manual. It also includes the reference to the used reference lists for the field recording to follow changes in the reference lists. The *existing data sources* reflect existing ongoing or historic monitoring activities. The data are structured according to the original requirements of the monitoring programme therefore are and are heterogeneous and in some cases semantically in-interoperable. There is need to structurally and semantically harmonise the existing information with the newly collected data to ensure proper analysis. This harmonisation is done in the *Data Integration Component*. This is either implemented as data warehouse solution (which can handle structural and semantic information) or using spatial web services for providing the data.

The *data presentation component* aims to present either raw or aggregated data or the resulting indicator values.

There will be the need for a *result database component* which contains the result of the external indicator calculation, if this is not directly implemented as analysis process in the data warehouse.

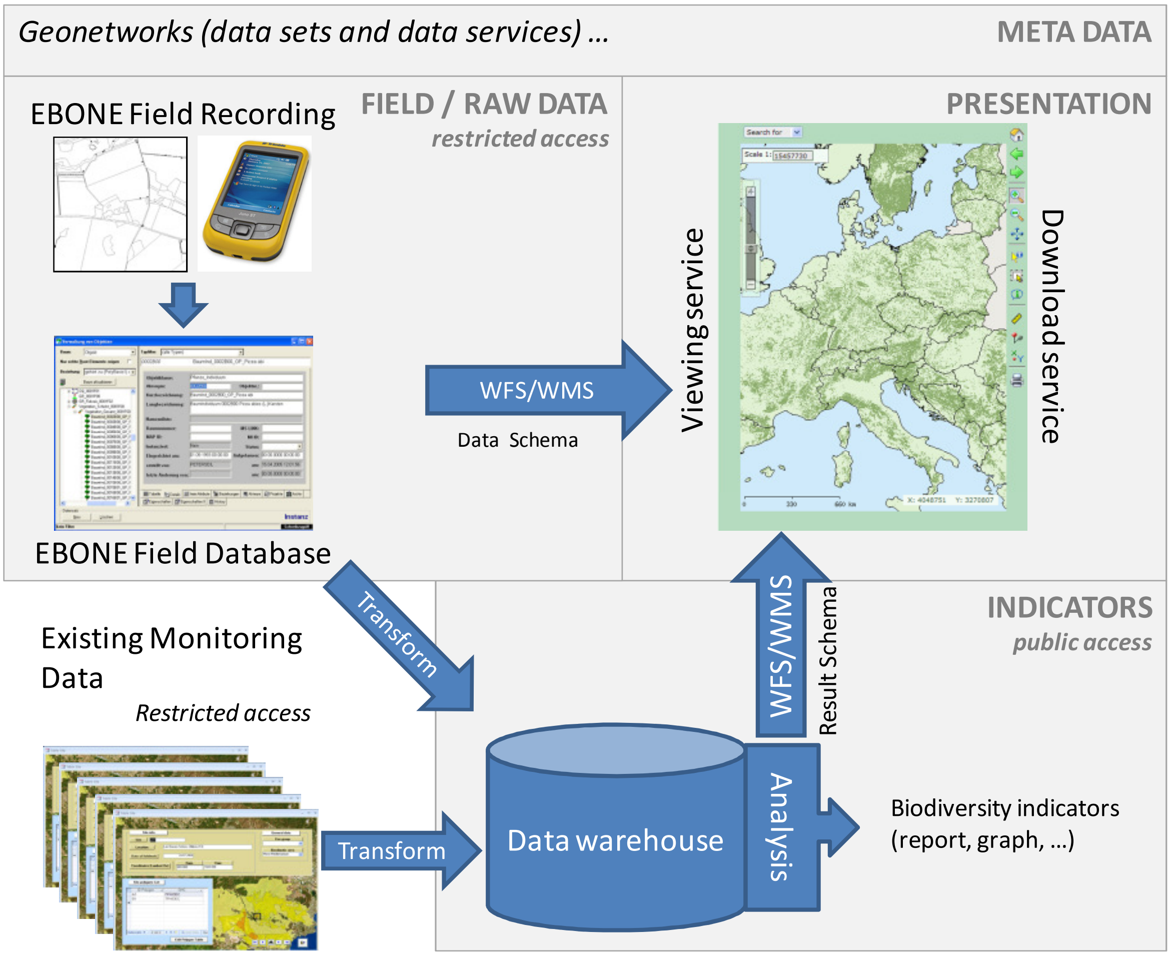


Fig. 3 Overview of the EBONE Data Management Components

The *Source Data Component* also need to include a central repository for the reference lists which were used in the field as well as the standard reference lists distributed data sources are translated to. This translation and harmonisation will be done in the *Data Integration Component* or hardcoded in the source data before integration.

The *metadata component* allows the notification of the relevant INSPIRE compliant metadata for the data sets and the data services. There will be the need for the further discussion of metadata fields dealing with the content of the dataset or data service.

5.2 Data flow

Based on this architecture the data flows within EBONE can be identified. In principal there is no difference in the process between the species and habitat information. The raw data are transformed according to the transformation rules defined by WP4 followed by harmonisation of the data. The results are structurally and semantically unified data on an aggregated level (see data levels) depicted in a common domain model for the habitat and species data. Therefore the common domain model is of eminent importance for the interoperability of data and components. The calculation of the indicator values or further scientific analysis must be based on harmonised data. The schema below (see Fig. 4) shows the reference model of the data flow, regardless of its technical implementation. The common domain model can either be implemented in central data storage or by a virtual central data storage using data ware house or semantic data integration technologies.

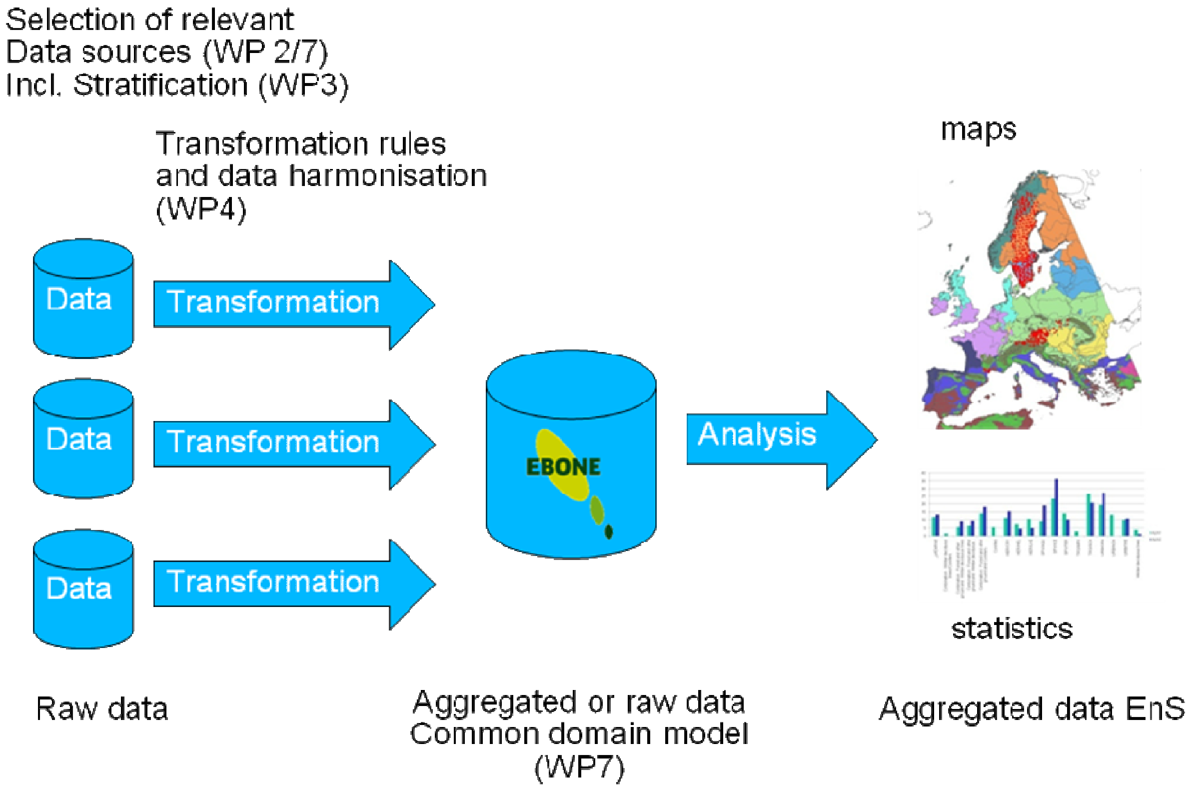


Fig. 4 Schema of the data flows for the EBONE data management

For the current implementation of the transformation and harmonisation of local reference lists, e.g. species lists, habitat lists, to the standard reference list (global list) will have to be done on the level of the data sources and/or ETL in order to provide harmonised data. Transformation of the data model will be done by the services (see Chapter 6.3.1.1) or during the ETL process (see Chapter 6.3.2).

5.3 Data model

As mentioned above, a common domain model needs to be established for the integration of existing as well as newly collected data. The use of standards and commonly used data models will at least on a certain level ensure the semantic interoperability, also beyond the current project. In the frame of INSPIRE and GEOSS the OGC Observation and Measurement (OGC, 2007) is one of the recommended standards (Nativi et al., 2009). Within the Network of Excellence ALTER-Net the Core Ontology SERONTO (van der Werf et al., 2009) was established which aims to form an integration core for environmental monitoring and biodiversity assessment and monitoring on the level of LTER sites and LTSER platforms.

In the following chapter the core concepts of the existing domain models are explained. In the further work EBONE will define the common domain model for habitat and species monitoring based on this core model. An application schema for the configuration of the spatial web services as well as the model for the data warehouse will be be derived from this.

An important prerequisite for the sharing of data and applications is the existence of structured metadata with common description formats and semantics. While most systems for the exchange of ecological data in operation today have provisions for attaching relevant metadata to the data, currently none of the widely used systems address the semantics of the information effectively. One of the most effective methods currently available for the representation and storage of such semantically enhanced metadata is the use of ontologies (Schleidt & Schentz 2006, Schentz et al. 2006).

An ontology is defined as a collection of concepts and their relations to one another based on logic (Sowa 2000) and established by a community with the purpose of sharing knowledge, information or data (Gruber 1993, Guarino & Giaretta 1995). Considering that "people can't share knowledge if they don't speak a common language" (Davenport 1997), the real challenge for developing ontologies lies exactly in finding this common language and understanding. This commitment to a common language seems to be essential for the acceptance and reuse of the ontology.

5.3.1 Definition of terms

In this chapter a short introduction to the main terms used in the EBONE data management is given. They reflect the current state of discussion and will change with the project. These concepts will be part of the common domain model.

Landscape square - a landscape square is a selected rectangular observation plot (usually with the extent of 1x1km according to the INSPIRE GRID or national grids) where the observation of the landscape structure, habitats and occurring species (vascular plant species) is done during the field mapping.

Landscape element - a single landscape element is a defined area in the landscape which is delineated according to the dominant habitat and the main ecological characteristics. Parameters (as GHC, qualifiers, etc.) are recorded for the single landscape element. Within EBONE the landscape element is also named as *polygon*.

Vegetation plot - a vegetation plot is a permanent monitoring structure which is revisited to describe the species composition of a specific landscape element.

Habitat – used according to the definition of the GHC (see WP4)

Species - in this terms vascular plant species

This requires the following fields in the data management:

DATA TABLES

- LANDSCAPE_SQUARE
  - SQUARE_ID - landscape square identifier
  - SQUARE_NAME - name or local name of the landscape square
  - SQUARE_AREA - area of the landscape square in km²
  - ENVIRON_ZONE_ID - assignment of the landscape square to an environmental zone (according to the table ENVIRONMENTAL_ZONE)
  - COUNTRY_ID - identifier of the country (according to the table country)

- LANDSCAPE_ELEMENT
  - SQUARE_ID - landscape square identifier (according to the table LANDSCAPE_SQUARE)
  - LANDELEM_ID - identifier of the landscape element
  - GHC_ID - identifier of the general habitat category (according to the reference list GENERAL_HABITAT_CATEGORY) including the version of the reference list used.
  - AREA_HA - area of the landscape element in ha (=100m²)
  - OBS_DATE - observation date in the format yyyy/mm/dd (e.g. 2010/01/12)
  - OBS_AUTHOR - author of the observation
  - optional additional qualifiers can be present from the original data source

- VEGETATION_PLOT
  - SQUARE_ID - landscape square identifier (according to the table LANDSCAPE_SQUARE)
  - LANDELEM_ID - identifier of the landscape element in which the vegetation plot is located (according to the table LANDSCAPE_ELEMENT)
  - VEGPLOT_ID - identifier of the vegetation plot
  - VASC_SPEC_ID - identifier of the vascular plant species (according to the table VASC_PLANT_SPECIES) including the version of the reference list used.
  - ABUND_VALUE - abundance value (e.g. 5)
  - ABUND_METH - method to estimate the abundance (e.g. Braun-Blanquet)
  - OBS_DATE - observation date in the format yyyy/mm/dd (e.g. 2010/01/12)

REFERENCE LISTS

- ENVIROMENTAL ZONE
    - ENVIRON_ZONE_ID - Identifier of the environmental zone
    - ENVIRON_ZONE_SHORTNAME - Short name or abbreviation of the environmental zone
    - ENVIRON_ZONE_NAME - Name of the environmental zone
    - NUM_OF_CELLS - number of square kilometer cells (according to the INSPIRE GRID) which represent the environmental zone in Europe
- GENERAL HABITAT CATEGORY (GHC)
    - GHC_ID - Identifier of the GHC
    - GHC_SHORTNAME - Short name or abbreviation of the GHC
    - GHC_NAME - Name of the GHC
    - EUNIS_HABITAT_ID - assignment of the most probable EUNIS Habitat category
- VASC_PLANT_SPECIES
    - VASC_SPEC_ID - Identifier of the vascular plant species
    - VASC_SPEC_SHORTNAME - Short name or abbreviation of the vascular plant species according to Ehrendorfer 1973, e.g. fag sylv
    - VASC_PLANT_NAME - scientific name of the vascular plant species, e.g. Fagus sylvatica L.
    - VASC_PLANT_EN - English name of the vascular plant species
- COUNTRY
    - COUNTRY_ID - identifier of the country
    - COUNTRY_SHORTNAME - short name or abbreviation of the country
    - COUNTRY_NAME - name of the country

Each of the original data sources need to be mapped to the common reference lists of species (SYNBIOSYS) and habitats (GHC). Harmonised data need to be transferred to and from the *central repository*. The transfer can either be done by an ETL (data warehouse) as well as a WFS service. In the implementation strategy both of the ways need to be described.

The following figure is not yet finalised but reflects the main components of the database which need to be implemented. A more detailed data model will be worked out after the workshop.

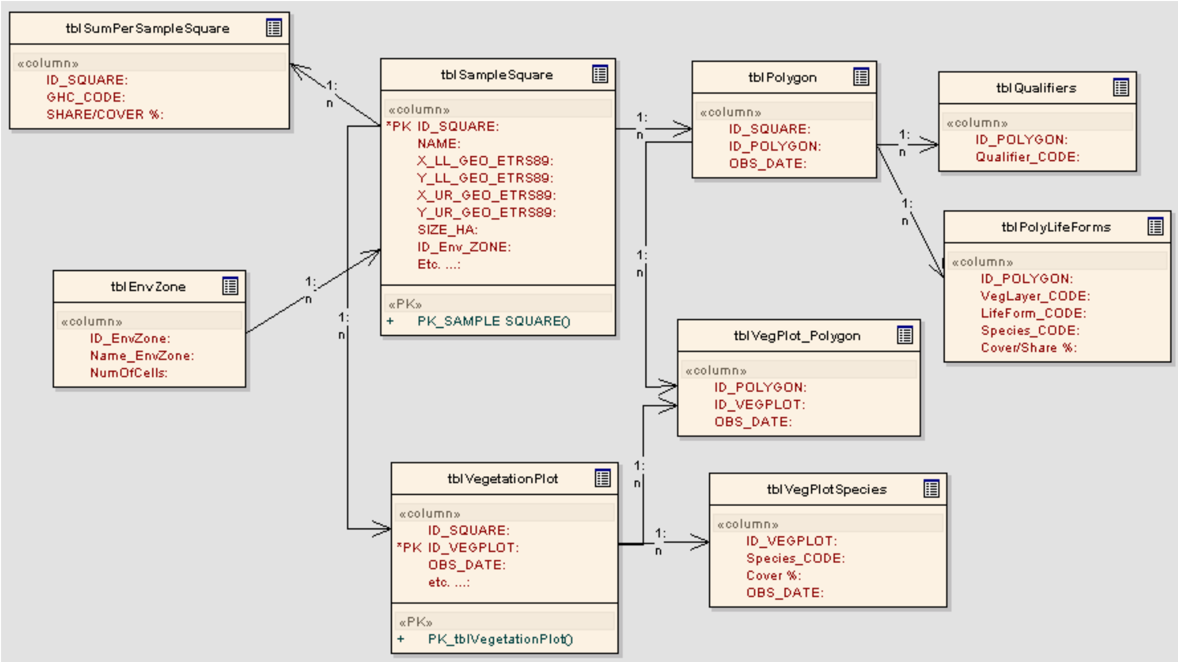


Fig. 5 Conceptual model for EBONE field observation based on the discussion within the data workshop 06/2009 Vienna

5.3.2 OGC Observation & Measurement

Observation and Measurement (O&M) has been issued by the Open Geospatial Consortium (OGC, 2007) and is in the process of being published as ISO standard. The development of the standard was supported by OGC through the OGC Web Services (OWS) Interoperability Projects, by the Water Resources Observation Network activity based at CSIRO Australia, and by GeoScience Australia managed primarily by Simon Cox. Observation and Measurement must be considered as one key component needed for the enablement of a "Sensor Web" through which applications and services will be able to access sensors of all types over the Web. The other components belonging to the same framework are: Sensor Model Language (SensorML), Transducer Markup Language (TML), Sensor Observation Service (SOS), Sensor Planning Service (SPS), Sensor Alert Service (SAS). O & M is primarily required for the SOS and related components of an OGC Sensor Web Enablement capability. Parts of the description of O & M use other components just like the procedure representation being based on SensorML. The model has been formalized in UML notation and XML Schema. It leverages standard components for features and geometry from other OGC standards. The Observation model takes a user-centric viewpoint, emphasizing the semantics of the feature of interest and its properties. Its documentation is split into two packages: one for the core model and one for a sampling feature model.

Model description:

The aim of the *core model* is to define a number of terms used for observations and measurements and the relationships between them. An *observation* serves as a property-value-provider for a feature of interest. In other words an observation is interpreted in O&M as an event with a discrete time instant or period through which a number, term or other symbol is assigned to a phenomenon. For the cases the result is a numeric quantity the whole model describes a measurement. The phenomenon is a property of an identifiable object, the so called *featureOfInterest* of the observation. The key properties of an observation are its *featureOfInterest*, the *observedProperty*, the procedure and the result.

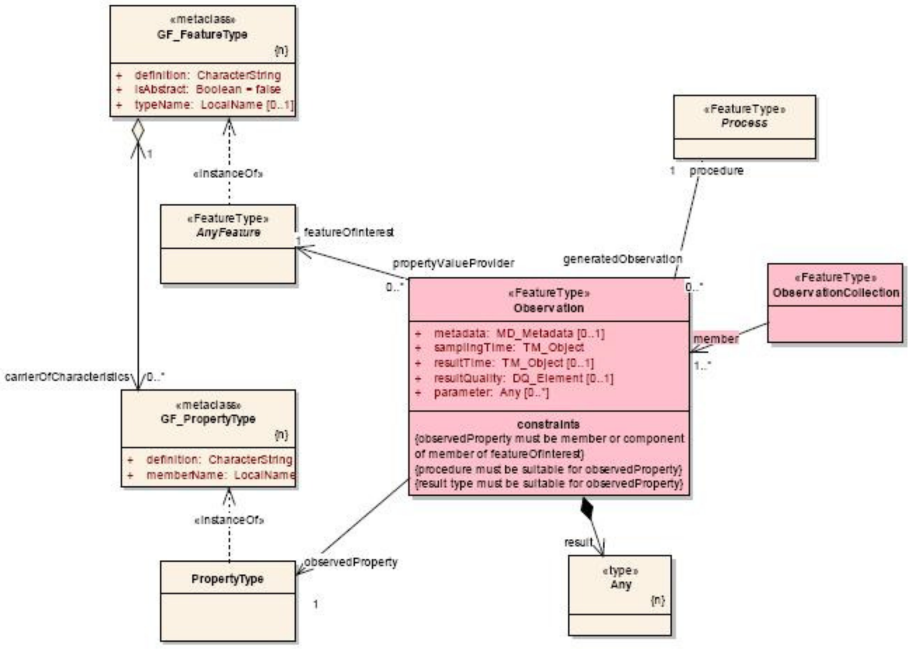


Fig. 6 Core model O&M

The *featureOfInterest* is a feature of any type being a representation of the observation target which is the real-world object observed. The feature type defines its set of properties, whose values could be determined either by some rule or assertation (e.g. name, ownership) or by observation.

The *observedProperty* identifies the phenomenon for which the observation result provides an estimate of its value. The feature property could hold a structure (e.g. feature associations) so in this case the observed property may be of one component of the complete feature property. If the type of feature allows for a property that is dependent on some parameter, then the value of the property is a function of this parameter. If the variation is temporal or spatial, then the function is a coverage whose domain is the spatio-temporal extent of the feature.

The *procedure* describes the process used to generate the result being suitable for the observed property. This could but not need to involve the use of sensors (or any other instruments), methods, algorithm, computation, simulation or systems of these or simply the human observer. The description of the process chain is done with SensorML, which is a very generic XML-specification allowing comprehensive description of complex processes.

The *result* is an estimate of the value of a property of the feature of interest generated by the procedure. The type of the observation result must be consistent with the observed property. The possible data types range from primitive types like category or measure to more complex types such as time, location and geometry (including coverages). The key idea is that the result is the primary output of the model whilst the other observation properties provide context or metadata to support evaluation (e.g. of errors in the estimate), interpretation and use of the result.

An observation may embrace a complex process over an extended period. Two time-related properties are distinguished in such cases: the *samplingTime*, being the time that the result applies to the feature-of-interest (e.g. when the specimen was retrieved from its host) and the *resultTime* being the time when the procedure associated with the observation act was applied (e.g. when the laboratory procedure was carried out). When this distinction is not necessary the *resultTime* may be omitted.

An observation *parameter* is a general event-specific parameter. This could be an environmental parameter or event-specific sampling parameter that is not tightly bound to either the feature-of-interest or the procedure. An observation may have additional *metadata* and an indication of the event-specific *resultQuality*.

Observations may be associated with a geospatial location, but the relevant location information should be provided by the feature of interest or by the observation procedure.

An *ObservationCollection* is composed of a set of member observations. There are principally three different types of collections: the *ComplexObservation* with same feature of interest, same sampling time but different observed properties; the *TimeSeriesObservation* with the same feature of interest, the same observed property but different sampling times, and the *DiscreteCoverageObservation* with the same observed property, the same time but with features of interest that comprise elements of a larger feature.

In many practical cases, observations are not performed on the feature of ultimate interest of an observation, either because the feature is inaccessible (in this case the focus lies on a subset of the complete feature, with the intention that the sample represents the whole) or because the properties are not directly observable (however, there are sensible properties that may be combined and/or further processed to obtain an estimate of the property of interest). These challenges are normally met by using proximate "sampling features" which are accessible and have properties that are sensible (proximate observed property). Sampling features embody a sampling strategy that is suitable for the observation procedure and the observed property. Similar sampling designs are used across a wide range of application domains which are described in the *sampling feature model* (see Fig. 6).

A *sampling feature* must be associated with one or more other features through an association role *sampledFeature*. This associations holds the intention of the sample design whose target is usually a domain feature. A samplingFeature has a set of navigable associations with Observations, given the role name *relatedObservation*. Sampling features are often related to each other, as parts of complexes, networks (such as sampling points on a sampling curve). This is documented by the *relatedSamplingFeature* association with a *SamplingFeatureRelation* association class, which carries a *source*, *target* and *role*. A *samplingFeatureCollection* has the specialized relation *member*. A *SurveyProcedure* provides the *surveyDetails* related to the specification of its location and shape. Different kinds of *SpatiallyExtensiveSamplingFeatures* based on their shape are *SamplingPoint*, *SamplingCurve*, *SamplingSurface* and *SamplingSolid*. *Specimen* is a physical sample usually carried out ex situ, often in a laboratory with properties like *currentLocation*, *size*, *materialClass*, *samplingTime*, *samplingMethod* and *processingDetails*.

The use of sampling features or sampling properties always involves to some extent the application of a transforming procedure to obtain an estimate of the ultimate observation target. An observation processing chain may require a series of transformations of the result coupled to transformation of the observed property. The consequence is often a change in the feature of interest between steps whereas the latter must always be consistent with the result, carrying the observed property as part of the definition of its type.

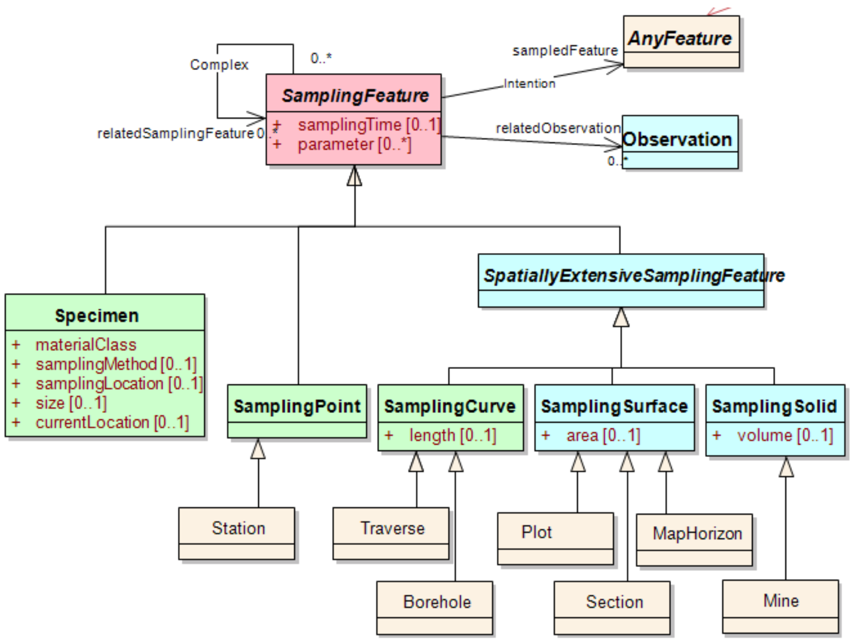


Fig. 7 Basic sampling feature model

O&M provides a generic structure for observation data. Domain specialization requires the development and management of vocabularies and ontologies to be used as values for elements within a data instance. These may include:

- Domain feature types (e.g. organism, organism occurrence, ecosystem)
- Observable properties (e.g. location, taxon, size, frequency)
- Observation procedures
- Scales and reference systems for observation results (including taxonomies)
- Sampling-feature relationship types (e.g. part-whole, manifolds, networks and topology)
- Specimen preparation procedures

There are already several possible ontologies and schemas containing the domain information needed. The specification document refers in this context to *SWEET*, one of the most well known ontology for physical properties, published by NASA. Based on the core model for observation and measurements *GeoSciML* is an UML model/application schema based on O&M that specifies a set of feature-types and supporting structures for information used in the solid-earth geosciences and is scoped approximately to the information required to construct geologic maps. It is primarily concerned with "interpreted" geology (units, structures, etc), but links to external schemas for the descriptions of observational data. *WaterML* is a standard XML schema for hydrological information also based on the core of O&M. This XML specification has already been transformed to OWL to build an ontology. OWL (Ontology Web Language)[2] is used within *WaterML* to develop vocabularies on top of O&M skeleton.

5.3.3 SERONTO

The ALTER-Net ontology SERONTO (Socio-Ecological Research and Observation oNTOlogy) consists of a core ontology[3] and a separate unit and dimensions ontology[4] which form together the base for all other ontologies build on top of them. The core provides descriptions for the most important aspects of data derived from monitoring, experiments and investigations thus enabling a seamless presentation of data from different origins in the same conceptual manner. The concepts of the core are derived from scientific principles and lean heavily on statistical methodology while adhering to W3C standards and INSPIRE principles. This nucleus is encircled by a common knowledge base which holds descriptions used by all other domain ontologies on top of them (e.g. ecosystems, biodiversity, taxonomy, geography, norms), which extend the concepts and relationships of these two inner layers for their specific needs and requirements (see Fig. 8). While the core has already been formalised in OWL-DL following the whole creation process including the decision procedure, the common knowledge space and the domain ontologies are still in process of being conceptualised. The compatibility to OBOE (SEEK Extensible Observation Ontology), EML (Ecological Metadata Language), and OGC Observation and Measurement is also aimed to be provided within the SERONTO framework.

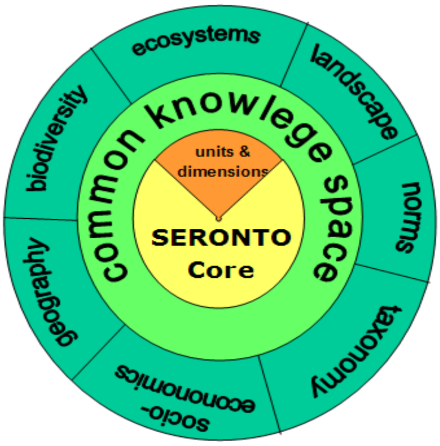


Fig. 8 Circular structure of SERONTO components

[2] The W3C standard ontology language

[3] See http://www.umweltbundesamt.at/fileadmin/site/daten/Ontologien/SERONTO/SERONTOCore20090205.owl

[4] See http://www.umweltbundesamt.at/fileadmin/site/daten/Ontologien/SERONTO/Units_Dimensions20080812.owl

Important considerations in designing SERONTO have been (van der Werf et al. 2008):

1. Repeatability: The ontology should be capable of holding enough meta-data that another person can repeat the experiment or observation at another place and time. It is not obligatory, however, to provide all information for all datasets; for instance, some information may be missing for old datasets.

2. Transparency: It must be possible to record and retrieve meta-data describing what actually happened. SERONTO includes concepts of things going wrong and documenting data collection under less than ideal conditions. If data and meta-data are available in this way, it will be clear what assumptions must be made to combine data and correctly interpret analyses.

Central concepts in the SERONTO core are (see general schema in Fig. 9):

1. Physical thing – in the role of an investigation object (can also be the experimental unit);

2. Parameters – the measurement, classification and treatment of the investigation object;

3. Value sets – joined concepts holding the information for the investigation object, the combined parameter/method used and the time series of values;

4. Reference elements – pointing to reference and reference lists such as species lists (necessary intermediate concept due to the fact that the same references could be part of different reference lists) at disposal for any used concept and specially for nominal values;

5. Methods – used for each parameter, including units, scale, and dimensions – could be also represented as method chains;

6. Selection descriptions – the origin of the research object or population including the sampling method;

7. Groupings of objects, such as experimental blocks, on which observer, time or other aspects are assigned or related to;

8. Additional information, such as actors (observer, observer groups and institutions), project information, etc., can be attached to several different concepts.

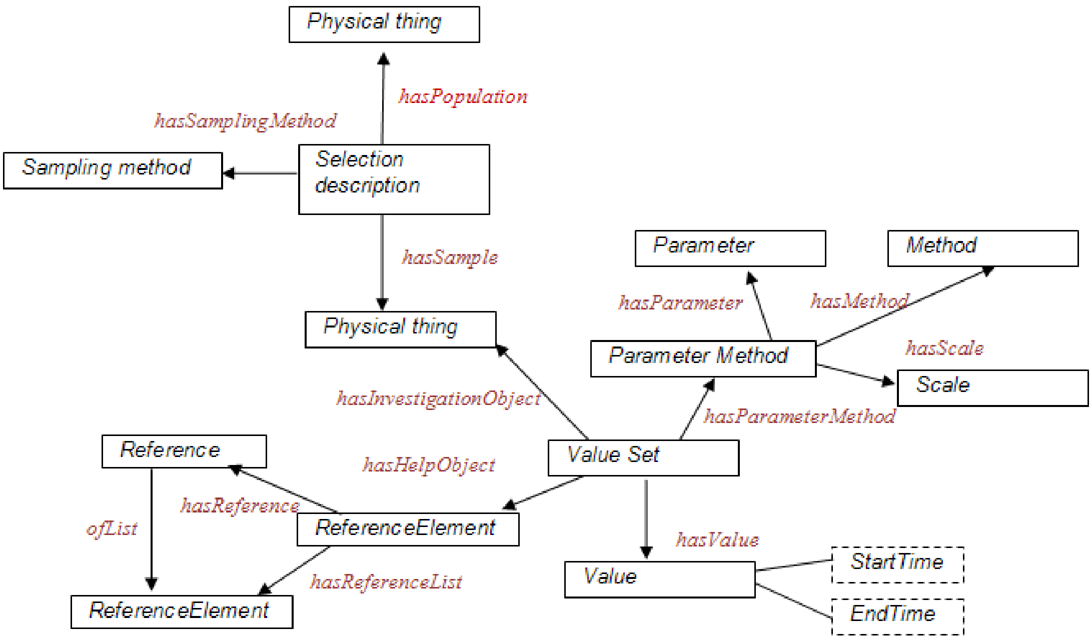


Fig. 9 General schema (classes) for an observation (SERONTOCore)

In comparison to OBOE[5] (Madin et al. 2009) the additional value of SERONTO seems to lie in the use of reference lists, exact descriptions of applied methods (allowing also chains of methods and methods encompassing other methods), the introduction of selection descriptions explaining the origin of the research object, the time stamp bound to every value and the provision of templates for specific domain use.

To prove the validity of SERONTO a quality check is foreseen to be performed in the next months. This will include a feasibility test of the ontology with various examples from ecological and socio-economic observation and measurement realities to be entered within a reasonable time. In addition also an integration test with real measurement data stored in relational databases should be undertaken to prove SERONTO's mapability. As next step a query ontology will be built to check the consistency of SERONTO which also implies the use of reasoner.

5.3.4 Implementation

SERONTO is hosted by LTER Europe and will further develop in this frame. Together with the emerging standards as the OGC Observation and Measurement it will be the basis for the EBONE data management system.

[5] See http://cvs.ecoinformatics.org/cvs/cvsweb.cgi/seek/projects/kr-sms/kr/ontologies/OBOE/

6 Implementation strategy and plan

The following chapter describes in short the technical solutions and the concepts used in the testing phase. This includes the components of the data management framework described in chapter 5.1.

The aim of the EBONE data management framework is to set up an architecture which is compliant to the INSPIRE directive and supports the GEOSS data sharing principles. The data management framework needs to be set up in a cost efficient manner. Therefore the use of standard and existing tools for the implementation is planned. The results of projects like GIGAS[6] or LifeWatch[7] are taken into account.

6.1 Metadata Management Component

One of the major challenges facing ecologists today is to obtain qualitatively and quantitatively sufficient described information pertaining to widespread ecological phenomena. The difficulties start when one tries to ascertain how data was collected and how it may be validly analyzed and integrated with other datasets.

Many initiatives exist to improve data structural and semantic interoperability of data. For geospatial information, the INSPIRE directive and associated initiatives such as GMES[8] or SEIS[9] aim to provide a service-oriented, standards-based framework. However, although these approaches deal with semantics they are limited in their ability to capture or exploit it for effective discovery and interoperability. In ecology, data integration has so far been carried out through consolidation of existing databases or XML-based (Briesen et al. 2002, McCartny & Jones 2002), the most prominent examples of schemata being EML, GML, Darwin Core or ABCD. Efforts on the global as well as on a national level have been undertaken to standardize metadata collection and description. They aim to collect information about existing data from monitoring and research. Metadata languages are being developed for different purposes, as natural history specimen data (e.g. Darwin Core (Darwin Core 2004), ABCD), long term monitoring data (e.g. EML[10] (Michener et al. 1997, Jones et al. 2001)) and for geospatial data on the European level (e.g. ISO191115 for the INSPIRE directive[11]). The range of details varies from information about the location of observation sites and a rough description of the data collected (see e.g. ILTER meta data collection[12]), the description of data sets for the exchange of data and metadata (EML approach of the US LTER, e.g. Michener et al. 1997) to the proposal to describe the observation itself in a semantic way (Adamescu et al. 2007).

For the well-founded interpretation and analysis of data, it is necessary to know various parameters pertaining to **"how, where, when, on what, by whom, …"** the data was collected. This information ranges from coordinates and owner of observed sites and individual plots over how they were selected and with what methods the data was measured to what sorts of transformations were performed with the data. This information becomes even more important when one wishes to use data generated to answer one specific question in order to find an answer to a different question. It is absolutely indispensable when one starts to combine data from different sources,

[6] http://www.thegigasforum.eu/project/project.html
[7] http://www.lifewatch.eu/
[8] http://www.gmes.info/
[9] http://ec.europa.eu/environment/seis/index.htm
[10] See http://knb.ecoinformatics.org/software/eml/
[11] See http://inspire.jrc.ec.europa.eu/
[12] See http://www.mexlter.org.mx/

regardless if the same type of data from multiple sites is being used or if different types of data pertaining to one site are being integrated (Schleidt and Schentz, 2006).

Currently, whenever such data integration tasks are carried out, much time is spent in communication with the originators of the various data sources. Some facets of this information can be retrieved, some can be speculated on, some are irretrievably lost, i.e. when the original investigator is no longer reachable or when basic facts initially thought to be self-evident are forgotten.

As collecting data is often one of the most expensive components of scientific research, it seems a great waste not to reuse information that has already been collected for one question in other ways. If the proper metadata has been appended to the data itself, this data can be reused after years or decades, either on its own or in combination with data from other sources. This can help to make scientific work more efficient, and thus more cost effective. It can also help scientists find answers to questions that would otherwise be too complex to generate all the required data for.

When contributing metadata, one should be aware of two dimensions. First, the metadata should be presented in such a way that it is useful/understandable for researchers who may be outside the sub-domain of the original data provider and who may be separated by time. It is important that long term data be documented so that the data is useful to a wide range of scientists. Since EBONE data may be used by researchers outside Europe and by researchers from many different disciplines, care must be taken to avoid descriptions that use narrow jargon.

6.1.1 Importance of extensibility

In contrast to many traditional metadata systems, e.g. Dublin Core, which is mainly used in Libraries, a metadata system for the structuring of ecological data must include mechanisms for extension. As science continues to grow and develop, introducing new concepts and methods on a regular basis. It must be possible to incorporate these new concepts and methods seamlessly into the existing system without changing the existing structures. On the one hand this process must be kept as simple as possible in order to avoid data not being included into the network because it cannot be represented adequately. On the other hand, this process must be fairly strict in order to avoid multiple descriptions of the same concept.

6.1.2 Types of Metadata

A coarse, often used structure of metadata are the categories: administrative, descriptive and structural. Although this structure has been taken from the library domain (apparently the only ones with a metaview to metadata), it is quite relevant for ecological purposes too. Although this differentiation is not necessary in order to generate and use metadata, we believe that it gives a good overview of the basic types of metadata required.

6.1.2.1 Administrative

Administrative metadata mostly pertains to ownership and intellectual property rights (IPR). This covers the following information:

- who created the data
- who currently owns the data
- who is responsible for the database the data is stored in
- what are the conditions for use of the data (access rights and access policy)

- how can the data be accessed (download service, link, etc.)

But it also includes metadata about the metadata.

- who has created the metadata

6.1.2.2 Descriptive

Descriptive metadata describes the actual parameters of data measurement or collection. This covers facts such as:

- what was measured
- where was it measured
- what units are used
- when was it measured (temporal coverage)
- how was it measured (using what method, device, limits of detection, …)
- on what was it measured
- why was it measured -> formal hypothesis

It further covers basic technical information such as:

- format of the data (i.e. how many bytes are used for a number; is a comma or a decimal point used for the representation of real numbers)
- type of the data (i.e. integer, real, LOV)
- when was the data stored
- is the data in it's original raw format or has it been processed in some way
- the language of textual data
- reference lists used

6.1.3 Structural / Semantic

Structural metadata describes how individual bits of data relate to each other. One important use of structural metadata is in defining time-series. It can also be valuable for deciding if different data sources pertain to the same geographic region.

The most important use is to describe the meaning of entities by their relation to other entities (e.g. species by their relations to morphological items)

6.1.4 Structure of Metadata

In its simplest form, metadata can be seen as a list of notes describing the data. While this list may be quite complete, it will be difficult for somebody else to find the relevant information. In order to know exactly how the data were assessed, one usually still needs to consult the data originator. This is time consuming and error-prone and raises the question a metadata system should be established when the data user still has to consult the data provider for proper use of the data. For automated data discovery and retrieval systems and for automatic modelling applications it is fairly useless because of the lack of unambiguous link between data and their machine readable description. Whereas a human being can still guess and find other methods to get it clear, a machine can only process exact linking's.

The current standard metadata annotation are XML files with the structure defined by xml schema files (XSD). The XML Schema defines the meaning of the tags, but does not define the meaning of the contents within those tags. The knowledge about that meaning has to be established outside the system and be distributed among the community, who wants to share data, which quickly can produce ambiguity.

6.1.5 Implementation

The key criteria for supporting flexible metadata applications are those of technical requirement, compliance with international standards, user friendly interface and availability of necessary functions for handling metadata records (Rajabifarad et al., 2009).

To create INSPIRE compliant metadata a metadata editor on the INSPIRE geo portal is available. It can be accessed by the address http://www.inspire-geoportal.eu/index.cfm/pageid/342. The metadata model is according to the Implementing Rules for INSPIRE Metadata (INSPIRE, 2007).

In addition to that for EBONE existing metadata editor will be tested. This is on the one side GeoNetwork[13] and possible extensions describing the data content. GeoNetwork opensource is a standards based, free and open source catalogue application to manage spatially referenced resources through the web. It provides metadata editing and search functions as well as an embedded interactive web map viewer. It is designed to enable access to geo-referenced databases, cartographic products and related metadata from a variety of sources (GeoNetwork Cummunity, 2007). GeoNetwork was jointly developed by FAO[14], WFP[15], UNEP[16] and UN-OCHA.

GeoNetwork OpenSource supports the ISO19115/19139 metadata standard for spatial metadata as well as FGDC and Dublin Core. Adaptions for the INSPIRE metadata model as well as for the description of the dataset content will be needed.

Morpho http://knb.ecoinformatics.org/morphoportal.jsp#download is a metadata editor that is designed to create EML compliant metadata.

Experiences of US LTER and the usage of EML for data descriptions will be included to extend the metadata schema if needed.

EML can be used to extend INSPIRE in the following areas:

- Taxonomic coverage (INSPIRE does not appear to have a taxonomic component).
- Keywords (INSPIRE's keywords lack the specificity that EBONE's data require).
- Physical data descriptions

6.2 Source Data Management Component

Two software solutions within the project frame developed will be used for the *field database component*. On the one side a field computer application which allows the collection of data in the field with a PDA or similar device. In addition a Microsoft Access Application which can be either used on a tablet PC in the field or as desktop application in the office is developed. Both solutions are specified and described in the following sections.

[13] http://geonetwork-opensource.org/
[14] http://www.fao.org/
[15] http://www.wfp.org/
[16] http://www.unep.org/

6.2.1 MONBEL - Field computer application (INBO)

The aim of the application is to support the field mapping with digital recording of the data. The mobile field mapping application includes the modules: a) mapping of an areal feature (including qualifiers and life forms), b) mapping of a linear feature (including qualifiers and life forms), c) mapping of point features (including qualifiers and life forms), and d) vegetation relevées. A GIS module is not included.

6.2.1.1 System Architecture

An overview of the general architecture of the application is given in Fig. 10. Data exchange is done via XML files which are structured and defined.

The reference lists are uploaded as XML files to the mobile application and stored in a SQL-CE database. The recordings of the single elements are stored as XML files and can be uploaded to a database. For the field database the structure is defined and the import routines according to the existing HabiStat database need to be implemented. With the installation all relevant reference lists will be provided.

The mobile field mapping application has the functionalities to manage the reference lists, to set up the database structure and to manage the fields in the application. A central repository for the reference lists (XML-files), the installation files and the handbook will be created at surfgroepen to distribute this files.

The structure of the XML-Files will be provided in the updated version of this report.

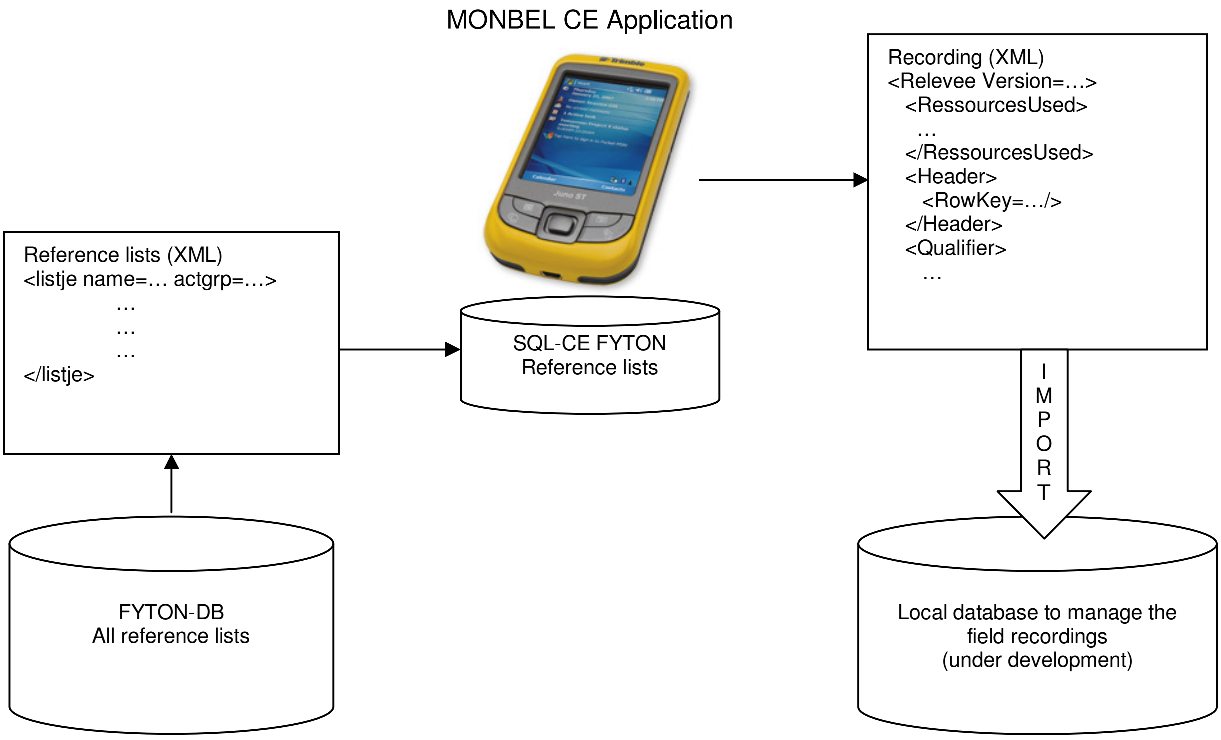


Fig. 10 Schematic architecture of the mobile mapping application

The application implements the protocol according to the EBONE field manual (Bunce et al. 2010). Different options for the recording can be chosen.

6.2.1.2 Application structure

Each relevé (recording, sample) is saved as a separate XML-file. The user is therefore free to import the xml-data in whatever database with whatever data model that can be matched by the data in the XML-file. At the PDA-level there is no hassle with a data model. The XML-files can travel very easily between different systems. A simple copy will do. And they are platform independent.

The recording-file also knows from which lists items were chosen. This is useful for diagnosing purposes: for example a given taxon (e.g. *Leptidea reali*, is a butterfly species) is never present in the samples. There could be two reasons for that: a) because the list at the time of sampling only names *L. sinapis*?, or b) because *L. reali* really wasn't there (assuming people on the field knew their job). With the link to the reference list you can get certainty about a).

The Application uses a SqlCE database for all reference lists (e.g. Taxa, Lifeforms, GHCs, Scales, Natura2000 habitats, etc.). Although lists can be loaded at will, this will typically only be done once at project startup and during PDA-preparation for the fieldwork and together with the project/research manager. The exchange of reference lists is also done as XML files.

The use of the reference lists for the fields of the application can be configured and changed. The link between fields and lists can be set independently for each sampling method, e.g. coverage of taxon: BioHab = percentage in 10% steps or Classic vegetation relevé = Braun-Blanquet or Tansley scale. The validation of the data (e.g. the sum of the coverages in a given layer) is also dependent of the chosen method.

A help section will be available to make configuration and list-load easier.

Validation rules, program configuration and -setup (not user configurable) are stored in XML files.

Furthermore the User can set many preferences to influence the behaviour of the application and to speed up his work. One of the more powerful options is the TaxonHistory which memorizes the x latest taxa which were recorded (even between sessions). A user at work in a given habitat will most likely encounter the same species again and again. The history makes it faster to pick the same taxon again.

There is also a mechanism present to cope with plants not known to the recorder. It is possible to make your own species (f.e. *Solanum* with spines (put in bag A25)) and use this description throughout the fieldwork. Back at the lab; the references can then be changed to the correct taxon before import to the database.

No spatial information is stored directly in the database. The spatial information is digitised using a GIS software (e.g. ArcGIS) and linked to the database entries using the identifier fields.

6.2.1.3 Technical Requirements

Minimum requirements

- Windows CE 4.2 and 5, PocketPC-2003, and Windows Mobile 5.x and 6.x

- Microsoft .Net Compact Framework 2 SP2
- Microsoft SQL Mobile 2005 (SqlCE 3.0)

The device needs an SD-card to write the recordings.

You'll also need tools to copy files (installation CABs; recordings) back and forth between your PC and the device. It's perhaps easiest to take Microsoft ActiveSync 4.2 (or above).

Microsoft Active Sync 4.2 (or above) must be installed on both the PC and the PDA.

Installation

Make sure that the PDA is connected to the PC (via a USB cable) and that Active Sync is up and running. Cancel the wizard that will offer synchronization.

Even the minimum installation requires that the so-called CAB files are copied from the PC on to the PDA. Use the Windows Explorer on the PC to transport the installation files to the PDA. In the Windows Explorer the PDA is indicated as 'Mobile device'. Remember the folder on the PDA to which the files are copied (e.g. 'Personal' or 'My documents').

1) Prerequisites

- Compact Framework 2, if not already installed (select the associated CAB file for your PDA)
- SqlServerCE3.0, if not already installed (select the associated CAB file for your PDA)

2) The actual Ebone software

- 1-IVCEInit_Setup.CAB
- 2-IVCEInputCfg_Setup.CAB
- 3-IVCEImport_Setup.CAB
- 4-IVCEOpnameD_Setup.CAB

6.2.1.4 Implementation

Within EBONE the field computer application was developed by INBO (Belgium) and will be used during the field season 2010 for test. The data are stored as single XML files with a clear identification and will be integrated into a common field observation database (see the sections below).

6.2.2 Field database CEMAGREF

A second version of the field database is developed by Philip Roche et al. (CEMAGREF) for the field data collection using a tablet PC or Laptop in the field. The structure of the application is according to the EBONE Field Manual (Bunce et al., 2010).

The following requirements were identified in the discussion at the data workshop (Vienna 06/2009). This is not a complete list, but a first discussion:

- EBONE field manual and survey forms

- Repeated observations need to be possible
- Vegetation relevé need to be stored
- Aggregated information on the level of the sample square need to be stored.

The application is implemented with Microsoft Access. Forms and data tables are stored in a Microsoft Access database. The application allows the entry of the field data and the management of the underlying reference lists.

The screens and work flow is optimised for the use of the database in the field. The database application can be also used for the data entry in the office after the field work.

The data can be exported to Microsoft Access tables, Microsoft Excel tables or text files.

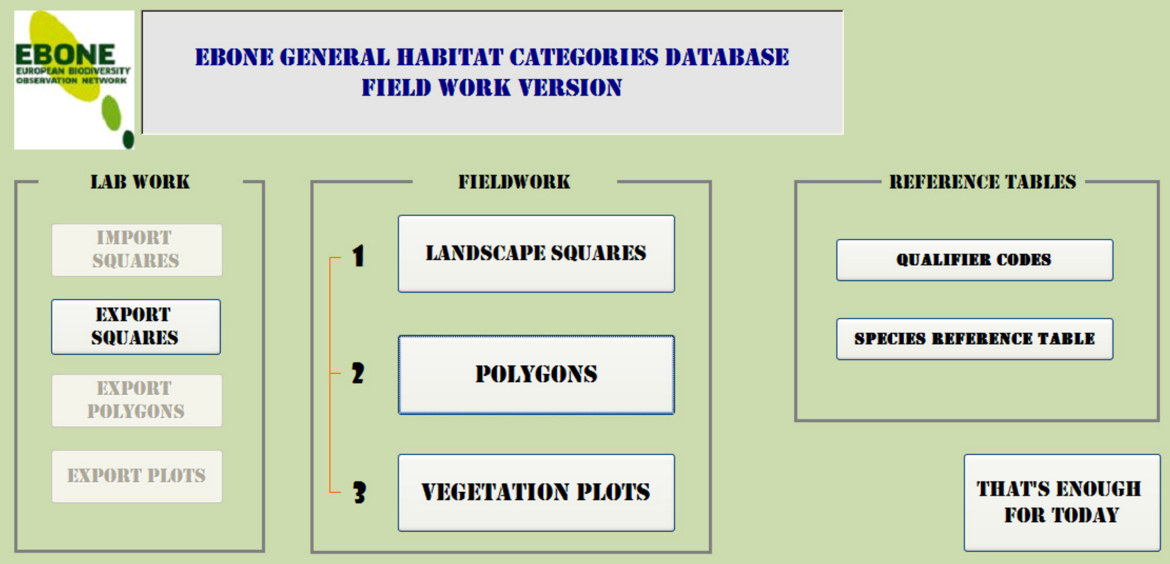


Fig. 11 Welcome Screen of the Field database application (CEMAGREF)

The user can enter data for the landscape squares, the polygons, and the vegetation plots (see Fig. 11). Appropriate forms for these levels of information are implemented.

6.2.2.1 Data model

The data model of the field database application (CEMAGREF) differs slightly from the data model from the MONBEL CE application (see chapter 6.2.1). The differences are mainly in the naming of the fields. A harmonisation the field names will be done after the field test season.

The main table areas of the database are:

- Site [Table_Site], which holds the data about the landscape square (e.g. name, location, date of mapping, etc.). This is the master table. A site is composed of 1-n polygons or landscape elements.

- Polygon [Table_Polygone], which holds the data about the mapped landscape elements according to the protocol of the EBONE field handbook. Sub tables containing the qualifiers and lifeform information are related 1:n to this.
- Plot [Table_Plot], which holds the data about the vegetation relevé.

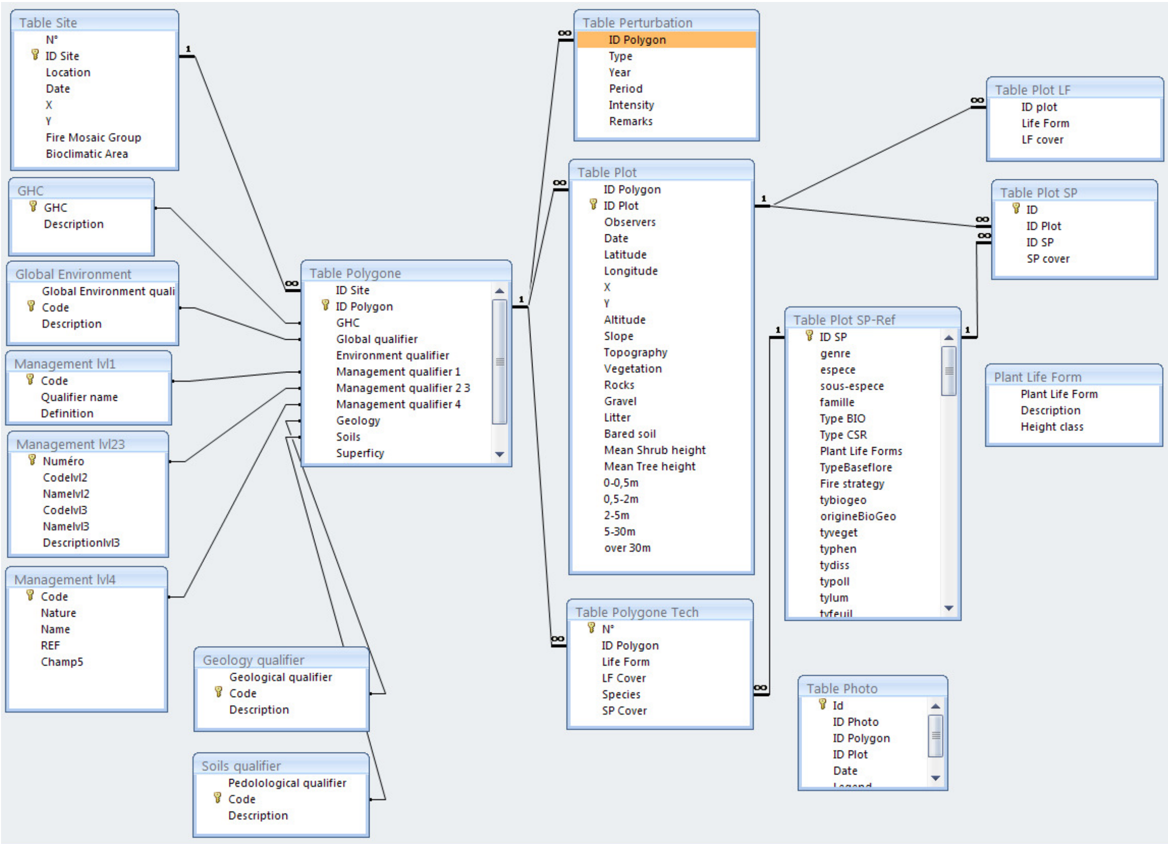


Fig. 12 Data model of the EBONE Field Database (CEMAGREF)

No spatial information is stored directly in the database. The spatial information is digitised using a GIS software (e.g. ArcGIS) and linked to the database entries using the identifier fields.

6.2.2.2 Application preview

In the following section some screenshots of the application are presented. Appropriate forms for the landscape square, the polygon, and the plot are implemented in the database.

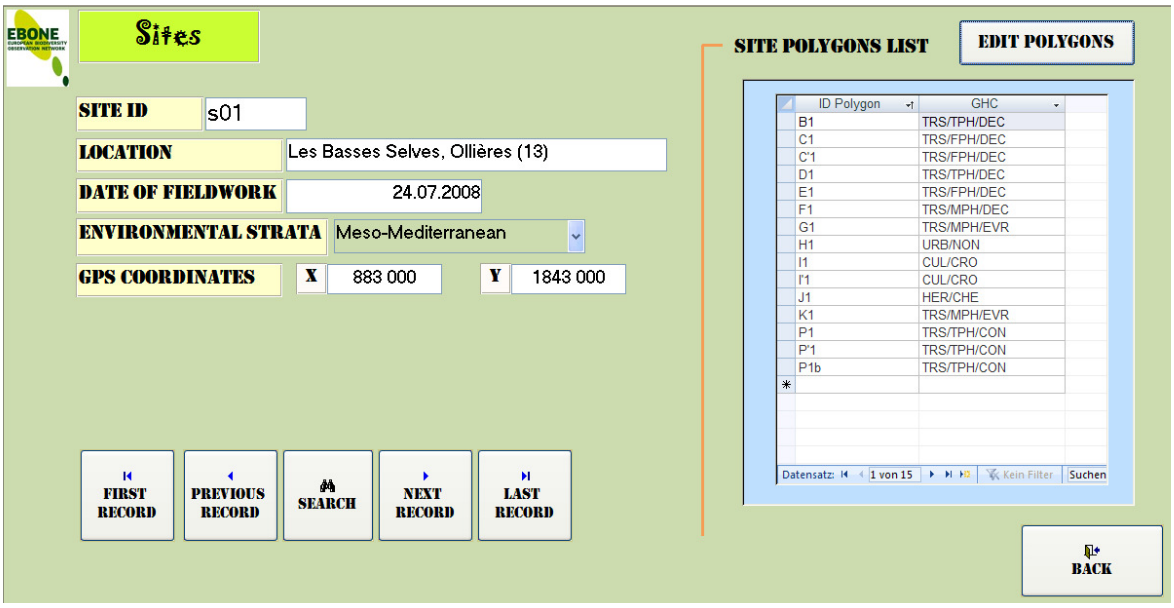


Fig. 13 Screen to enter landscape square data

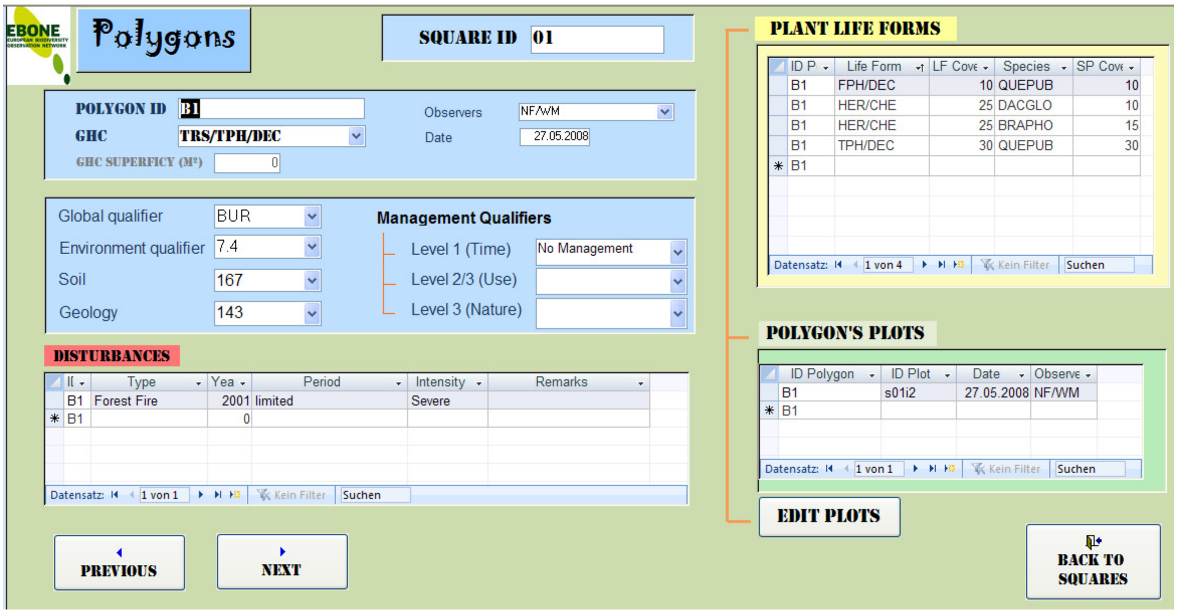


Fig. 14 Screen to enter polygon data

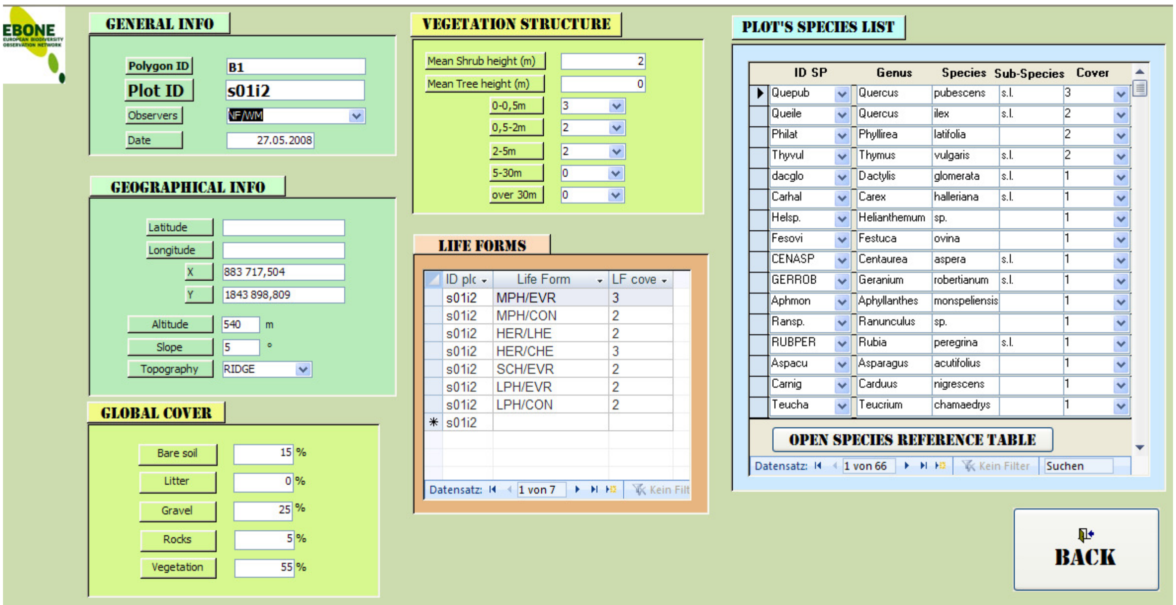


Fig. 15 Screen to enter plot data

6.2.2.3 Implementation

A Microsoft Access application for the use in the field as well as desktop data entry tool was developed by CEMAGREF (France) and will be used during the field season 2010 for the field tests. All field records on the national level will be collected in one database following the structure of the field database. This will be one of the data sources to be integrated and provide a harmonised view for the data analysis.

6.2.3 Analysis result data

For data resulting from evaluations and calculation data another database will be used. As this data are highly aggregated and the structure of the data is not known at the moment a generic database which is able to adapt to the needed data structure will be necessary. For this purpose MORIS (Monitoring and Research Information System, see Schentz et al. 2005, Mirtl & Schentz 1997, Schentz & Mirtl 2003) will be used. MORIS was developed at the Umweltbundesamt GmbH (Austria). MORIS is a object relational database which was designed to store heterogeneous structured data resulting from long term ecosystem research.

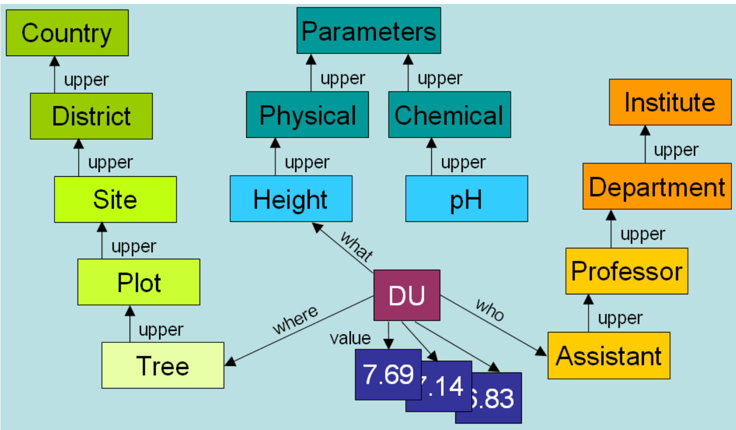
These data are characterised by varying spatial and temporal scale of observation and varying media and changing observation targets and methods.. They need to be integrated and made available for a long term storing and offering all metadata for correct interpretation. Those requirements were solved by a simple object oriented basic data model following the basic principles of ontologies.. The user is able to structure the domain data model based on the underlying core data model according to his needs, at runtime without any adaption of the user interface.

The challenge being faced today is the definition of a widely accepted standardized core ontology, a common basic vocabulary,with basic concepts and extensions for specialized needs as domain ontologies. MORIS provides the possibility to store thematic as well as spatial data.

MORIS provides a desktop client for the data management and a web client for data query. Both tools can be used within the runtime of EBONE.

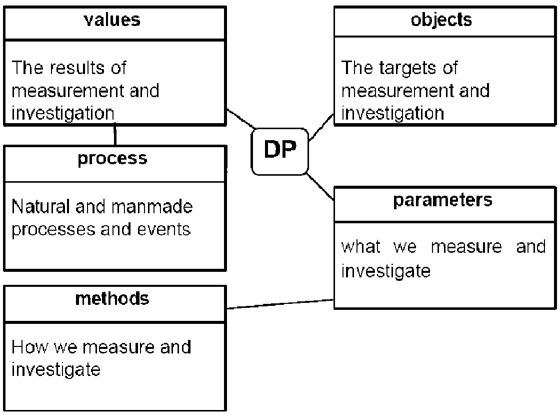
6.2.3.1 Data model

The core task of the MORIS is to facilitate the storage, administration and discovery of data on observations and measurements. **What** was measured on which **object** at which **location** using which **methodology** by which **person** at a specific **time**. Most of these concepts required for the description of an individual measurement or time series are usually standardized and can be neatly organized within one (or more) hierarchy.



This is done by using a simple core model to structure the information to be managed. In the current version of MORIS the modules object (target of measurement), parameter (characteristic measured), method (method of measurement), data unit (assignment entity), value (result of measurement) and specimen/process (specimen taken for the measurement) are used.

Within each of the modules the enduser defines classes, their relations to other classes (e.g. also a class hierarchy) and enters or imports instances for the defined classes. Thus he establishes a network of information around the results(values) of observations.



For the further development of MORIS a review of the concepts and the core data model into a widely agreed core ontology is done.

MORIS uses the basic model of "OBJEKT", "PARAMETER", "METHOD", "DATENPUNKT" and "VALUE" to structure and manage every environmental related information in an object oriented manner.

- The "OBJEKT" reflects the entity which is observed or is used as a reference in an observation. In the discussion of the ALTER-Net ontology (proposed name SERONTO) the concept might be renamed to "TARGET", "ENTITY OF OBSERVATION" or "EXPERIMENTAL UNIT". The "site/platform" and the "experimental unit" of the metadata collection are within this class, as well as the reference list entities of habitat types, species or soil types.
- The "PARAMETER" reflects the property observed or measured. This is directly related to the meta information in the LTER InfoBase MetaData Editor. Every parameter refers to a "METHOD" as the way how the parameter was measured or estimated.
- The "DATENPUNKT" brings together the entity observed with the parameter estimated. On the Datenpunkt the observed values are linked by an 1:n relationship.

For every end user defined class, derived from those basic classes, specific attributes and special relations to other classes can be generated.

Based on this simple basic data model, together with the enduser defined extensions nearly any ecological data can be stored in a structured way and managed by MORIS and relations between them are used to depict the information to be stored. A more extensive description can be found in Schentz et al. 2005 and http://www.umweltbundesamt.at/umweltschutz/oekosystem/informationssystem/.

The MORIS class model is in general compliant with the class model of OWL ontologies apart from minor exceptions. Unfortunately there is no OWL interface to expose models defined within MORIS in the web. However transformations with simple EXCEL Macros have already been done, to close this gap.

Currently MORIS is further developed as web based information system. Enhancement of the user friendliness and the core data model are the main points of the new version.

6.2.3.2 Implementation

The results from the indicator calculation will be stored in a MORIS instance. The data will be on the level of the Environmental strata and/or zones. Services to access those data within the EBONE architecture will be established during the test phase in 2010.

6.3 Data Integration Component

Within the EBONE data management framework different data sources are ,described in the *data source component* are integrated. Although they deal with the same domain of habituat and species monitoring, the different data sources show different data models There is the need to seemlessly integrate them for the analysis, download, and presentation.

The common domain model defined within the project will be used to map the different local models to and present them in a consistent structure and semantics . for interoperability reasons the common domain model will be based on OGC Observation and Measurement and SERONTO (see chapter 5.3).

Two different ways for data integration will be tested for the EBONE Data Management Framework. First the integration of existing data sources by data warehouse technology and second the use of a INSPIRE compliant web services to allow view and access to the raw data as well as to the aggregated data.

The following section gives an overview of the intended technology and an outlook how the presented tools and technology will be used during the testing phase.

6.3.1 Web services

Web services are web accessible applications and application components that exchange data, share tasks, and automate processes over the Internet. Because they are based on simple and non-proprietary standards, web services make it possible for computer programs to communicate directly with one another and exchange data regardless of location, processing platforms, operating systems, or languages (OGC, 2004).

The web service architecture distinguishes between three roles within the network: a) service provider, b) service requestor, and c) service broker – they perform three essential kind of operations – *publish, find, and bind*. The *Service Providers* publish machine-readable information – service metadata – about their service capabilities (as Web sites currently publish metadata about their data offerings). *Service Requestors* send out requests that announce what kind of service is requested. *Service Brokers* (relying on service catalogues and service registries) function something like today's search engines, receiving service requests and "binding" a service to a service request. After the service available from the Service Provider has been "bound" to the Service Requestor, the service executes. Services can be chained to create more complex applications. In the client-server model, the Service Requestor is a client, the Service Provider is a server, and the Service Broker is middleware (OGC, 2004; see Fig. 16).

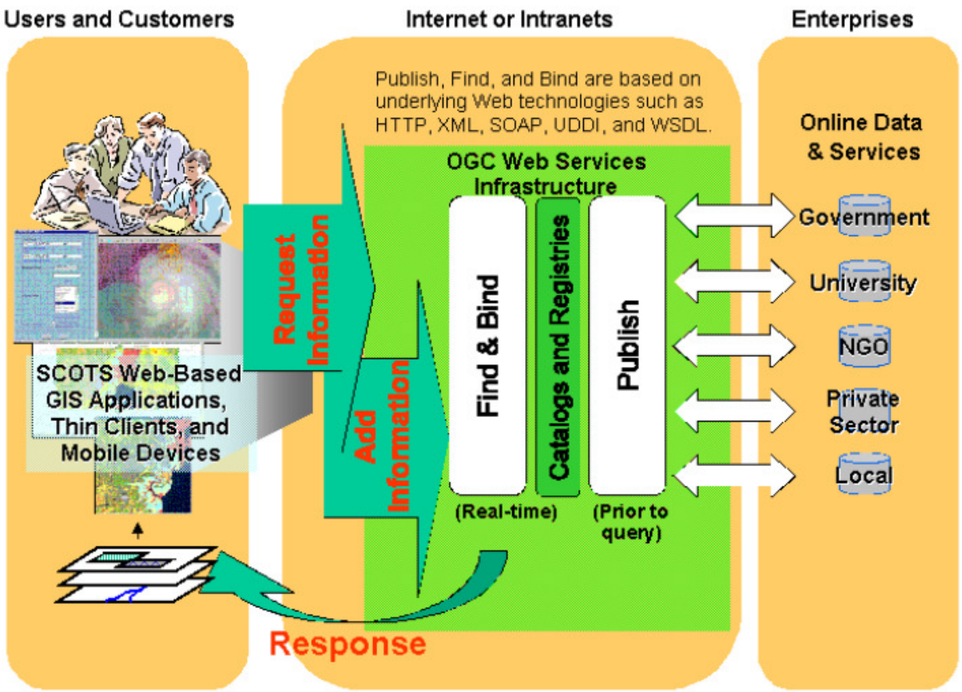


Fig. 16 Conceptual Archtitecture for OGC Web Services (OGC 2003)

6.3.1.1 WFS – Web feature service

The OGC Web Map Service allows a client to overlay map images for display served from multiple Web Map Services on the Internet. In a similar fashion, the OGC Web Feature Service allows a client to retrieve and update geospatial data encoded in Geography Markup Language (GML) from multiple Web Feature Services.

The requirements for a Web Feature Service are:

1. The interfaces must be defined in XML.
2. GML must be used to express features within the interface.
3. At a minimum a WFS must be able to present features using GML.
4. The predicate or filter language will be defined in XML and be derived from CQL as defined in the OpenGIS Catalogue Interface Implementation Specification.
5. The datastore used to store geographic features should be opaque to client applications and their only view of the data should be through the WFS interface.
6. The use of a subset of XPath expressions for referencing properties.

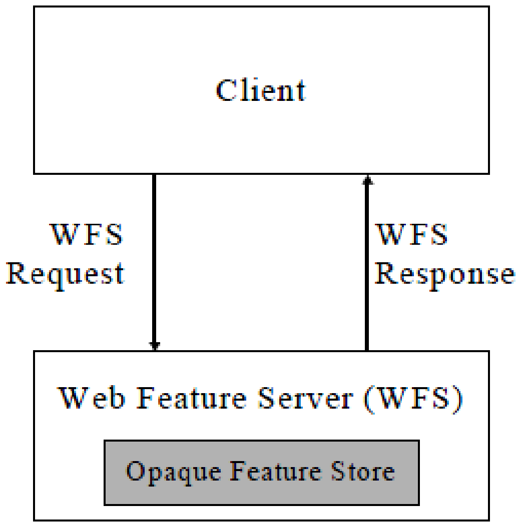


Fig. 17 Web Feature service (OGC, 2005)

TheWeb Feature Service (WFS) is a standard created by the OGC that refers to the sending and receiving of geospatial data through HTTP. WFS encode and transfer information in Geography Markup Language (GML), expressed in XML. The current version of WFS is 1.1.0. GeoServer supports both version 1.1.0 (the default since GeoServer 1.6.0) and version 1.0.0. There are differences between these two formats, some more subtle than others, and this will be noted where differences arise. The current version of WFS is 1.1. WFS version 1.0 is still used in places, and we will note where there are differences. However, the syntax will often remain the same.

An important distinction must be made between WFS and Web Map Service, which refers to the sending and receiving of geographic information after it has been rendered as a digital image.

WFS can perform the following operations:

Operation	Description
GetCapabilities	Retrieves a list of the server's data, as well as valid WFS operations and parameters
DescribeFeatureType	Retrieves information and attributes about a particular dataset
GetFeature	Retrieves the actual data, including geometry and attribute values
LockFeature	Prevents a feature type from being edited
Transaction	Edits existing feature types by creating, updating, and deleting.
GetGMLObject	(Version 1.1.0 only) - Retrieves element instances by traversing XLinks that refer to their XML IDs.

The benefits of WFS are that one can think of WFS as the "source code" to the maps that one would ordinarily view (via WMS). WFS lead to greater transparency and openness in mapping applications. Instead of merely being able to look at a picture of the map, as the provider wants the user to see, the power is in the hands of the user to determine how to visualize (style) the raw geographic and attribute data. The data can also be downloaded, further analyzed, and combined with other data. The transactional capabilities of WFS allow for collaborative mapping applications. In short, WFS is what enables open spatial data.

In principal simple features are transferred using a WFS service. For more complex data models the extension to application schema, which is implemented in GeoServer, could be used. In the use case of habitat and species monitoring this would be the hierarchy of landscape square and landscape element. Complex features contain properties that can contain further nested properties to arbitrary depth. In particular, complex features can contain properties that are other complex features. Complex features can be used to represent information not as an XML view of a single table, but as a collection of related objects of different types.

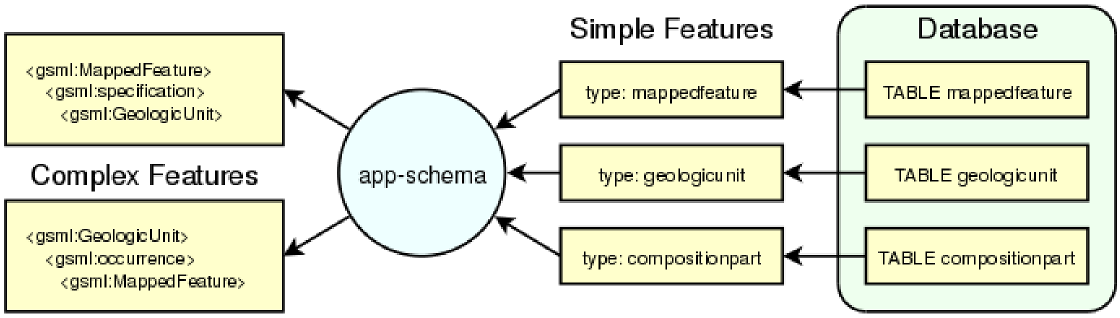


Fig. 18 Complex feature types using the application shema[17]

6.3.1.2 WMS – Web map service

The Web Map Service (WMS) is a standard created by the OGC that refers to the sending and receiving of georeferenced images over HTTP. These images can be produced from both vector and raster data formats. The most widely used version of WMS is 1.1.1, which GeoServer supports. The Styled Layer Descriptor (SLD) standard specifies extensions to WMS to control the styling of the WMS over the web.

An important distinction must be made between WMS and Web Feature Service, which refers to the sending and receiving of raw geographic information, before it has been rendered as a digital image.

WMS can perform the following operations:

Operation	Description
GetCapabilities	Retrieves a list of the server's data, as well as valid WMS operations and parameters
GetMap	Retrieves the image requested by the client
GetFeatureInfo (optional)	Retrieves the actual data, including geometry and attribute values, for a pixel location
DescribeLayer (optional)	Indicates the WFS or WCS to retrieve additional information about the layer.
GetLegendGraphic (optional)	General mechanism for retrieving generated legend symbols

The benefits of WMS are that it provides a standard interface for how to request a geospatial image. The main benefit of this is that clients can request images from multiple servers, and then combine them in to one view for the user. The standard guarantees that these images can all be overlaid on one another as they actually would be in reality. Numerous servers and clients support WMS.

[17] See http://docs.geoserver.org/trunk/en/user/data/app-schema/

6.3.1.3 Implementation

For the EBONE data management framework the use of WMS and WFS using GeoServer tools is planned. During the test phase 2010 services to external data sources (e.g. SINUS, NILS) will be established.

A view service will be implemented by using WMS and a download service by using WFS. The underlying data model specification will be derived from the common domain model.

The services will be established for selected external (= existing monitoring schema) and internal data sources (=newly generated monitoring schema). The services will be established either directly to ESRI shape files containing the spatial information but also directly to field database containing thematic as well as spatial information.

GeoServer is the reference implementation for the OGC Web service specification and provides support for Open Geospatial Consortium (OGC) Web Feature Service (WFS) versions 1.0 and 1.1. This is a standard for getting raw vector data - the 'source code' of the map - over the web. Using compliant WFS makes it possible for clients to query the data structure and the actual data. Advanced WFS operations also enable editing and locking of the data. GeoServer is the reference implementation of both the 1.0 and 1.1 versions of the standard, completely implementing every part of the protocol. This includes the Basic operations of *GetCapabilities*, *DescribeFeatureType* and *GetFeature*, as well as the more advanced Transaction, *LockFeature* and *GetGmlObject* operations. GeoServer's WFS also is integrated with GeoServer's Security system, to limit access to data and transactions. It also supports a wide variety of WFS output formats, to make the raw data more widely available. GeoServer additionally supports a special 'versioning' protocol in an extension: WFS Versioning. This is not yet a part of the WFS specification, but is written to be compatible, extending it to provide a history of edits, differences between edits, and a rollback operation to take things to a previous state (GeoServer, 2010).

GeoServer provides support for Open Geospatial Consortium (OGC) Web Map Service (WMS) version 1.1.1. This is a standard for generating maps on the web - it is how all the visual mapping that GeoServer does is produced. Using a compliant WMS makes it possible for clients to overlay maps from several different sources in a seamless way. The GeoServer implementation fully supports most every part of the standard, and is certified compliant against the OGC's test suite. It includes a wide variety of rendering and labeling options, and is WMS Servers for both raster and vector data. The WMS implementation of GeoServer also supports re-projection in to any reference system in the EPSG database, and it is also possible to add additional projections if the Well Known Text is known. It also fully supports the Styled Layer Descriptor (SLD) standard, and indeed uses SLD files as its native rendering rules (GeoServer, 2010).

6.3.2 Data warehouse

A data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis (Inmon, 1995). This definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve metadata[18].

Data warehousing is a collection of methods, techniques, and tools used to support knowledge workers to conduct data analysis that help with performing decision-making processes and improving information resources (Golfarelli & Rizzi, 2009). In a data warehouse no new information is required to be added; rather existing information needs to be rearranged. Data are never deleted from data warehouses and updates are normally carried out when data warehouses are offline. This means a data warehouse can be understood as a read-only database. One of the most obvious features of a data warehouse relational implementation is that table normalisation can be given up to partially denormalize tables and improve performance.

The data warehouse consists of the subsequent data flow stages (Lechtenbörger, 2001): a) source layer, b) data staging, c) data warehouse, and d) analysis. The source layer reflects the heterogeneous data sources. The data staging extract, cleansed to remove inconsistencies and fill gaps, and integrates to merge heterogeneous sources into one common schema. This is done by an Extraction, Transformation, and Loading Tool (ETL). This stage deals with problems that are typical for distributed information systems, such as inconsistent data management and incompatible data structures (Zhuge et al., 1996)

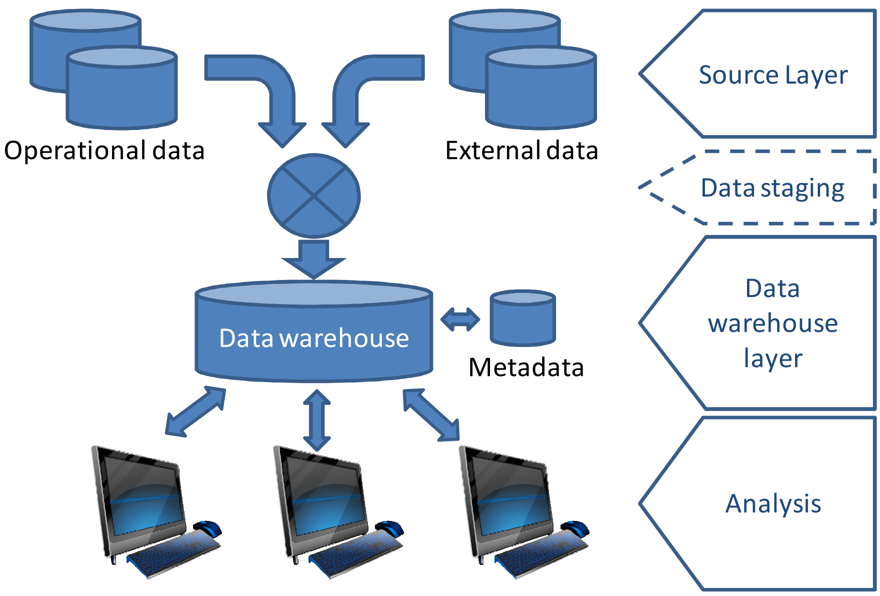


Fig. 19 Schematic architecture for a 2-layer data warehouse system (after Golfarelli & Rizzi, 2009)

One of the basic feature of a data warehouse is the data staging. The data staging layer hosts the ETL processes (*Extract-Transform-Load*) that extract, integrate, and clean data from a operational data source to feed the data warehouse layer. I a three layer architecture, ETL processes feed the reconciled data layer. ETL takes place once when a data warehouse is populated for the first time, thin it occurs every time the data warehouse is regularly updated. The ETL consists of four separate phases: *extraction, cleansing, transformation*, and *loading* (see Fig. 20).

[18] Source Wikipedia

During the *extraction phase* relevant data is obtained from the sources. This can be either static (e.g. when the data warehouse is first populated) or incremental (when the data warehouse is updated). The *cleansing phase* is supposed to improve data quality. This includes e.g. the removal of duplicate data, missing data, or inconsistent values. The *transformation phase* converts data from its operational source format into a specific data warehouse format. This phase is depended on the number of different heterogeneous data sources. The main transformation processes are *normalisation* and *conversion, matching*, and *selection*. When populating a data warehouse, normalisation is replaced by renormalization because data warehouse data are typically denormalised in order to enhance and optimise the query and analysis processes (see e.g. Golfarelli & Rizzi, 2009). The *load phase* loads the data into the data warehouse and is often carried out in two ways: *refresh*, which completely rewrites the content of the data warehouse, or *update*, which only updates records in the data warehouse which have been changed.

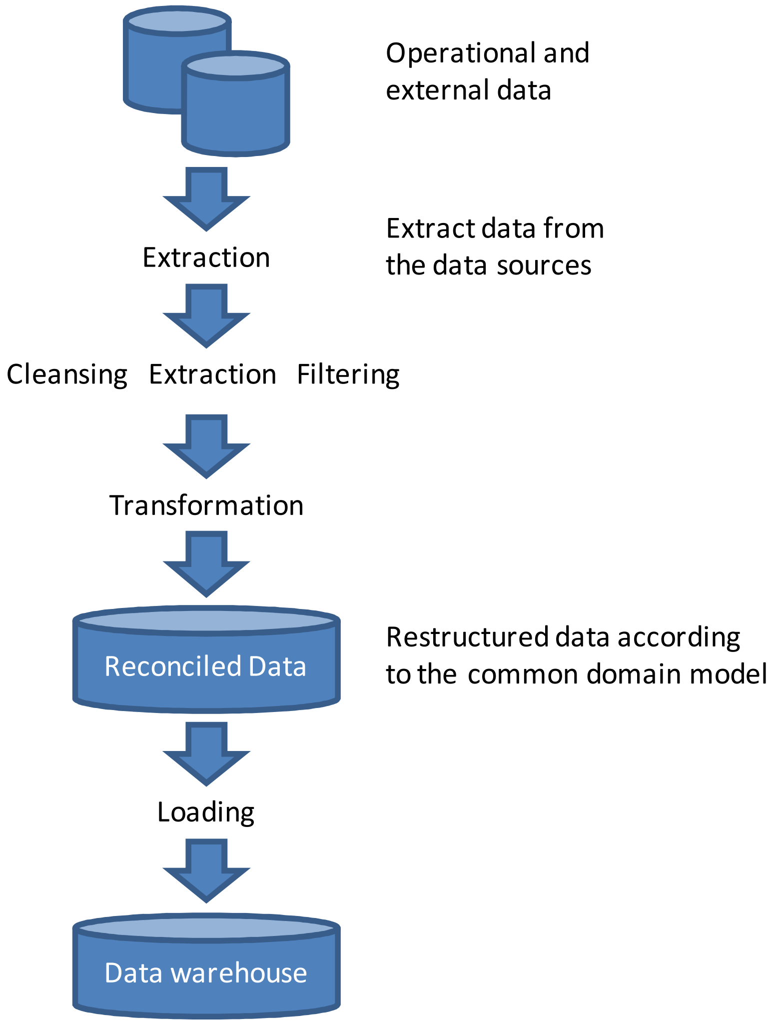


Fig. 20 Phases of the ETL process (after Golfarelli & Rizzi, 2009)

6.3.2.1 Implementation

Within the current project the open source data warehouse PENTAHO will be tested to establish a common database for the analysis based on the existing and newly created EBONE data.

The data model for the data warehouse will be derived from the common domain data model developed within the EBONE project. This test phase includes the following steps to be taken in the test phase 2010:

- Development of the common data model
- Set up the ETL process for the selected test data sources
- Execute the ETL process to populate the data warehouse

The content of the data warehouse reflects the harmonised data from all the originating data sources. The data warehouse will be used to populate the analysis for the biodiversity indicators.

In a later stage an update of the data warehouse will be made if the data are changed.

6.3.3 Implementation of the common domain model

Once the common schema is designed, the question of its implementation and use in the Web services set up arises. The common domain model is set up to provide a harmonized view and access to the data sources datasets through the Web Services, whereas the underlying data sources remain with their own coordinate system, structure and semantics. This architecture is one of INSPIRE principles: leave the data where and how they are and access them through distributed services. As a result, translation between the national models and the common schema is needed.

For the transformation there are thus two solutions:

- The transformation is done within the data source to comply with the common model". The transformation is stored in the database or as a query and hard coded way to solve the problem of transformation.
- The user's query is translated in a query compatible with the national data model. This is known as "on-the-fly" translation

During the reply process, similar situations occur:

Once the system has been able to answer the user's query, it sends to the requesting Web Service data in GML format. Here again, two solutions exist:

- The data has been transformed to comply with the common model. In that case, no translation is needed.
- Transformation is computed on the fly. In that case, the GML file must be transformed to a file compliant with the common model.

For EBONE the first option will be used. The transformation will be hard coded in the data sources. Therefore only harmonised data will be sent.

For the transformation the following reference lists will be used:

- General habitat categories according to the EBONE Field Handbook (Bunce et al., 2010)

- EUNIS habitat categories[19]
- Vascular plant species according to the Flora Europea

6.4 Data Presentation Component

The data presentation component describes the map viewer application which can be used as portal. For this portal the content is provided as web map service (WMS). Currently those kinds of web portals on the European level are evolving. One of the is the interactive map for EUNIS habitats at the web portal of the European Environment Agency[20]. For the topics forest and soil similar initiatives are implemented with the European Soil Data Centre[21] and the European Forest Data Centre (EFDAC)[22] hosted at the JRC.

The existing data centres will be used as a model for defining the concept to contribute to a European Biodiversity Data Centre from the project EBONE. The EFDAC model and tools will be the first starting point for the testing phase 2010.

6.4.1 Context

The aim of this section is to present forest data (particularly the forest pattern and fragmentation indicators) management framework within the European Forest Data Centre (EFDAC) and discuss the potential use of EFDAC data management architecture for EBONE. The focus will be given on the EFDAC Map Viewer application which is the EFDAC mapping subsystem designed to access and analyze the forest related maps stored at JRC. Nevertheless the EFDAC Map Viewer has its own particularities, specific design and differences, the general architecture, design model and overall functionalities could be used as the example for the development of EBONE data management framework.

The EFDAC hosted by the Joint Research Centre (JRC) of the European Commission has been established to supply European Union decision-makers with processed, quality checked and timely policy relevant forest data and information within the EU. The EFDAC in long-term should bring the added value integrating and building bridges to the decision makers and the society. The recently established EFDAC provides a gateway to data holdings and information on forest resources in Europe. In addition, the EFDAC supports the generation of value-added forest indicators on sustainable forest management on the basis of data collected by the Member States. The EFDAC resides in meta-data and data servers outside JRC; therefore the EFDAC technical platform is targeted to be interoperable with such external nodes for the exchange of meta-data and data. The framework for such interoperability is set by the INSPIRE Directive. The EFDAC is built on the basis of existing systems, such as the European Forest Fire Information System (EFFIS), the Forest Focus database, the European Forest Information and Communication Platform (EFICP) as well as integrates tools and applications developed by JRC (EFDAC Map Viewer, EFDAC Metadata Catalogue and it's management system, European Forest Resources Map Viewer, application handling forest dominant tree species distribution and current / future tree habitat suitability maps, etc.).

[19] See http://eunis.eea.europa.eu/habitats.jsp
[20] See http://eunis.eea.europa.eu/gis-tool.jsp
[21] See http://eusoils.jrc.ec.europa.eu/library/esdac/esdac.html
[22] See http://efdac.jrc.ec.europa.eu/

The EFDAC Map Viewer[23] is the mapping subsystem of EFDAC designed to support the access to the forest related maps stored at JRC. The maps that are currently handled by the service are grouped into the categories forest pattern and forest condition.

In the first part of this section the general design and the data management architecture of the EFDAC Map Viewer will be described. Later the functionalities as well as the user and data management requirements will be presented. At the end the several use cases according to the different data presentation options and the brief description of the integration of the Map Viewer into EFDAC through the metadata catalogue will be provided.

6.4.2 Application (EFDAC Map Viewer)

As it was mentioned before the EFDAC Map Viewer is the mapping subsystem of EFDAC designed to support the access to the forest related maps stored at JRC. The generated maps can be viewed and navigated either as original raster maps or as thematic maps derived aggregating the original data at different administrative levels (from NUTS 0 to NUTS 3). In addition to map navigation, the application has a number of functions such as querying, selection, gazetteer, and auto-identify which will be described later in this section.

Raw Data. The data flow of EFDAC Map Viewer is presented bellow (schema No 1). The raw data for this application is CORINE Land Cover and Forest / Non Forest mask. The raw data for forest condition maps is derived from Forest Focus database Level I plots managed by separate Forest Focus platform accessible only by authorized persons.

- ***CORINE Land Cover.*** The European-wide harmonized CORINE Land Cover datasets for years 1990 and 2000 (CLC1990, CLC2000) are based on high resolution Landsat imagery. They are the European wide data source at a rather fine scale (25 ha minimum mapping unit), informing on forest but also on agricultural and artificial surfaces, despite their limitations due to the forest definition and the mapping methodology. The CORINE Land cover data (CLC) is the ready-to-use, validated, multi-temporal, consistent and harmonized land cover data available at a relatively fine scale over the European territory and the last decades (25 minimum mapping unit and 100 m spatial resolution, and roughly for the years 1990 and 2000). The change analysis covered 21 European countries while for year 2000, United Kingdom, Sweden, Finland, Greece, Malta, Cyprus were also available (27 countries). In the future the European-wide multi-temporal land cover maps derived from Landsat TM (25 m) or from MODIS (300m) will be better alternatives.
- ***Forest / Non - Forest mask.*** For producing the Forest / Non-Forest map the fully automatic image processing methodology has been developed and applied. In order to avoid problems linked to phenological differences between images and related problems such as equalizing the radiometric content of all images, the processing was based on a scene by scene approach. It was used the Pan-European Forest / Non Forest Map with target year 2000 (Data Source: Landsat ETM+ and Corine Land Cover 2000, Classes: forest, non-forest, clouds/snow, no data; Method: automatic classification performed with an in-house algorithm; spatial resolution: 25m). More about Forest / Non Forest mask http://forest.jrc.ec.europa.eu/forest-mapping/forest-cover-map/2000

[23] See http://efdac.jrc.ec.europa.eu/viewer

- ***Forest Focus.*** The raw data for forest condition maps derived from Forest Focus database Level I plots (systematic network of observations points arranged on a grid throughout Europe at a spacing of approximately 16km x 16km). The original Forest Focus data is managed by Forest Focus Platform which is accessible only to authorized persons therefore in the EFDAC Map Viewer only aggregated data as the vector maps are available.

Processing. It should be noted that no processing is being performed within EFDAC Map Viewer application. All processing has been done outside application and the results have been put in the database which serves the Map Viewer. As it was mentioned above the application handles raster maps and the vector thematic maps derived aggregating the original data at different administrative levels. The processed raster maps are stored in Oracle database as geotiff files. This forest spatial pattern raster map was generated from the forest/non forest binary Corine Land Cover data. It provides at pixel-level the spatial distribution of seven forest pattern classes. The processing was done by GUIDOS[24] software which is the Morphological Spatial Pattern Analysis software customized sequence of mathematical morphological operators targeted at the description of the geometry and connectivity of the image components. The foreground area of a binary image was divided into the seven generic MSPA classes Core, Islet, Perforation, Edge, Loop, Bridge, and Branch.

The vector maps have been processed based on three methods and GIS techniques. Data inputs were forest-non forest masks, the forest spatial pattern maps obtained by applying the mathematical morphology based software GUIDOS, the landscape patterns maps obtained by applying the landscape mosaic index and the equivalent connectivity area index derived from the Conefor Sensinode software. The analysis was conducted to demonstrate the methods with the only readily available, harmonized, relatively fine-grained and bi-temporal European-wide land cover data from CORINE Land Cover (100 m spatial resolution, 25 ha minimum mapping unit) of years 1990 and 2000.

Aggregation. The aggregation is being performed based on selected criteria - aggregation level (from NUTS 0 to NUTS 3). When the user selects one of available aggregation level the application aggregates the data and displays it to user.

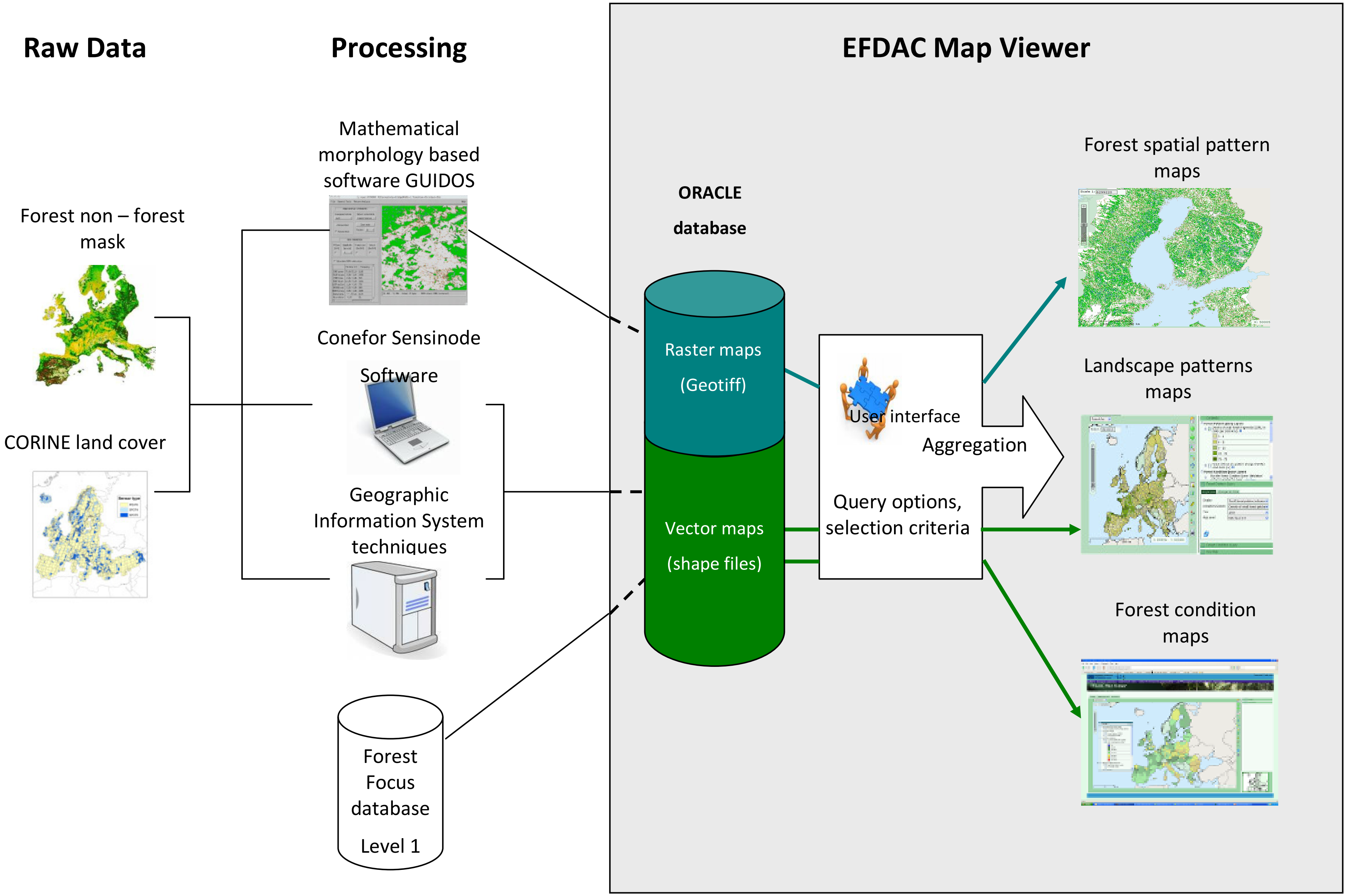
For forest patterns vector maps the local spatial information could be aggregated per province (NUTS level 2 or 3, - 564 provinces in total) and results could be presented on the basis of European-wide maps and tabular data.

For forest condition vector maps the aggregation level can be either countries or NUTS (Nomenclature of Territorial Units for Statistics) level 1 or 2. Depending on the year, the extent of the map may vary.

The data flow schema for the EFDAC Map Viewer is presented bellow (schema No 1).

[24] See http://forest.jrc.ec.europa.eu/download/software/guidos

Schema No 1 "EFDAC data flow"

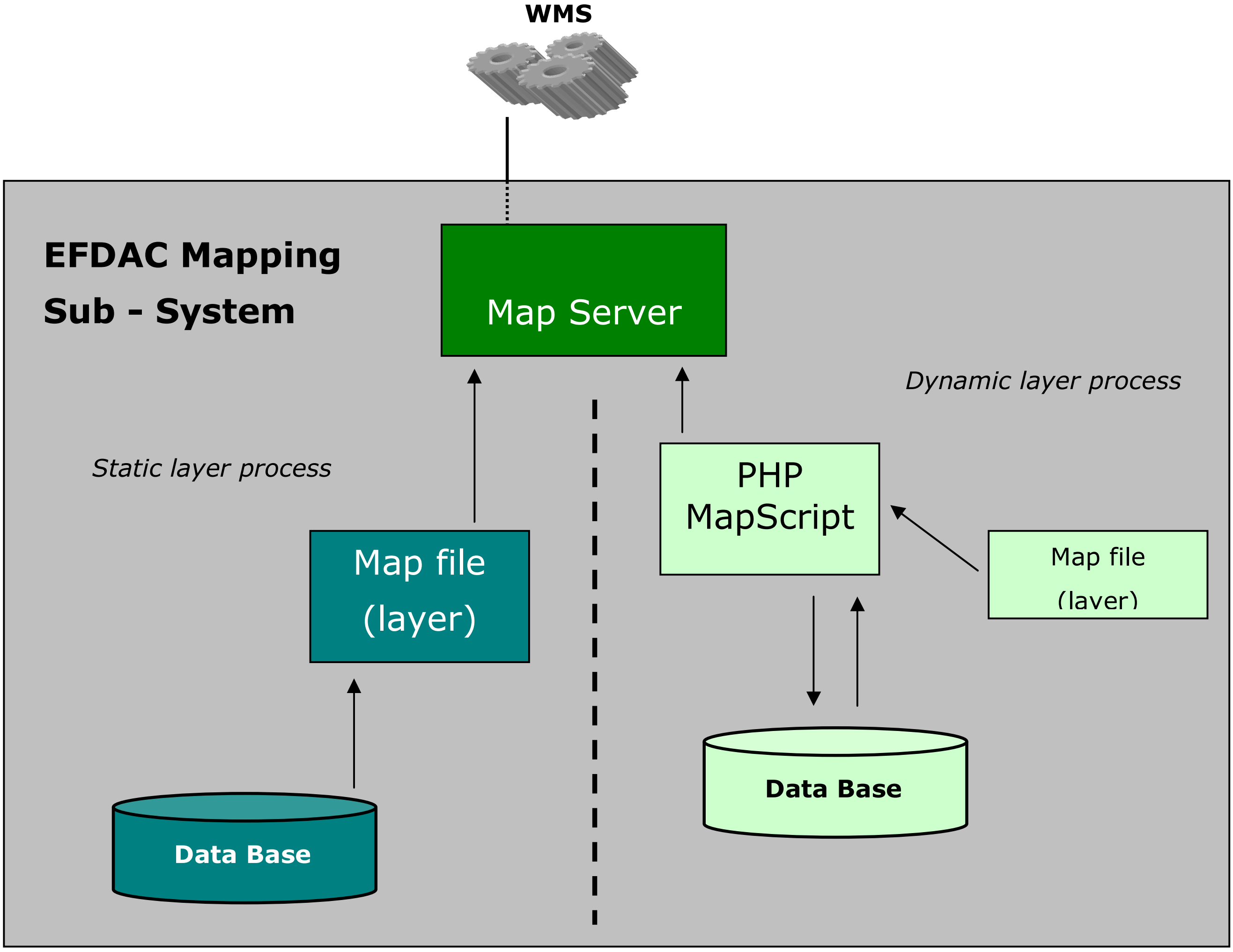


6.4.3 Software, interface

- ***EFDAC.*** The EFDAC Map Viewer is implemented using the p.mapper framework (http://www.pmapper.net/) and the MapServer (http://mapserver.org/) application which is based on PHP/MapScript. Vector datasets are stored as shape files in the Oracle database and raster datasets - as geotiff in the Oracle Spatial database. The EFDAC map layers are exposed through WMS service (OGC/INSPIRE compliant) as the selected (static) set of layers that are specific to this map viewer (see schema No 2, "Static layer process"). In addition, EFDAC Map Viewer can access trough WMS service the European Soil Data Centre (ESDAC) and display the pre-selected ESDAC layers as the static maps. This was designed to show that layers from different data sources could be displayed in the same Map Viewer.
- ***Other or/and additional options.*** Designing the EBONE data management framework, particularly the Map Viewer, the twofold approach could be explored: 1) aggregation of WMS/WFS services (and corresponding map layers) existing in other identified systems; and 2) publishing other mapping information that eventually doesn't exist in other mapping systems. For aggregation of WMS/WFS services existing in other systems, in addition to the static layer process, the dynamic layer process could be used (see schema 2 "Dynamic layer process), therefore the mechanism in order to create dynamically a selected set of layers that are specific to the map viewer (for instance: select a range of values per parameters and per layer) could be developed. This approach could be useful in EBONE data management framework when, either pre-created or dynamically created layers could be seen as regular WMS layers following the OGC WMS protocol and therefore interoperable with compliant clients. The EBONE data management framework could be designed in such a way which would allow to display any map information served by WMS services and also to allow download datasets via WFS services. These services can be accessed via results of metadata search mechanisms (Catalogue viewer) or adhoc mapping services by direct usage of their URL addresses.

Nevertheless the EFDAC Map Viewer is implemented using the p.mapper framework and the MapServer application, in the long-term it is planned to shift this application to the open – source software packages. Therefore EBONE should consider the possibility to use open - source software. In such case the baseline software packages and tools could be, for instance GeoNetworks (http://geonetwork-opensource.org/) and GeoServer (http://www.GeoServer.org) / UMN MapServer (http://mapserver.org/) packages. To serve as (geospatial) repository, metadata and data, the PostgreSQL, PostGIS or Oracle packages could, respectively, be used supporting both GeoNetworks and GeoServer/MapServer packages. The OpenLayers (http://www.openlayers.org) package could be used to implement the EBONE Map Viewer (and dataset download mechanism). Therefore, designing the EBONE data management framework all mentioned possibilities should be explored.

Schema No 2 "WMS

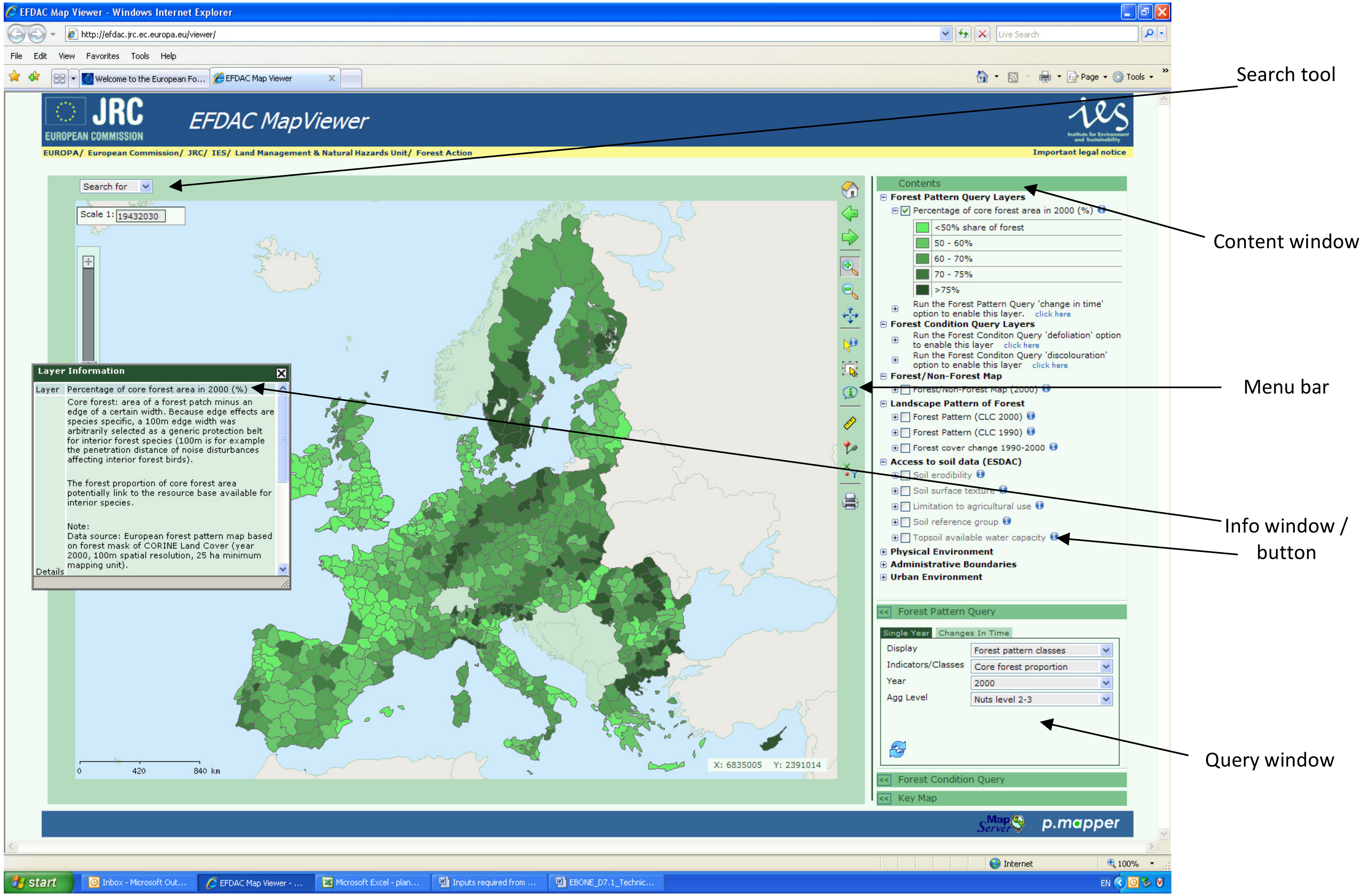


Functionalities. As it was mentioned before the EFDAC Map Viewer besides the map navigation, has a number of functions such as querying, selection, gazetteer, auto-identify, etc. The purpose of the EFDAC Map Viewer application is to present the raster maps as well as vector maps on forest pattern and fragmentation indicators and forest conditions in the user friendly interface allowing the user to query available maps using query options and different criteria as well as to display and navigate them for various research purposes. The functions are described bellow (see Schema No 3).

Description of functions. In the content windows (see Schema No 3) the list of available raster maps as well currently queried vector maps (of the current session) for every theme are displayed. The user can select / deselect the maps which allow him to manipulate and quickly change the layer on the screen. It is very useful for various analyses, in case the user wants to look for a different fragmentation indicator map in the same area of interest and quickly shift to other layer without performing the query once again. The legend of the map is shown bellow the layer title. In order to get additional information about the layer user can click the info button besides the title of the layer which opens the small info window with detailed information about the current layer. In addition, the user can manipulate (select /

deselect) the Physical Environment, Administrative Boundaries and Urban Environment layers.

Schema No 3 "EFDAC Functions"

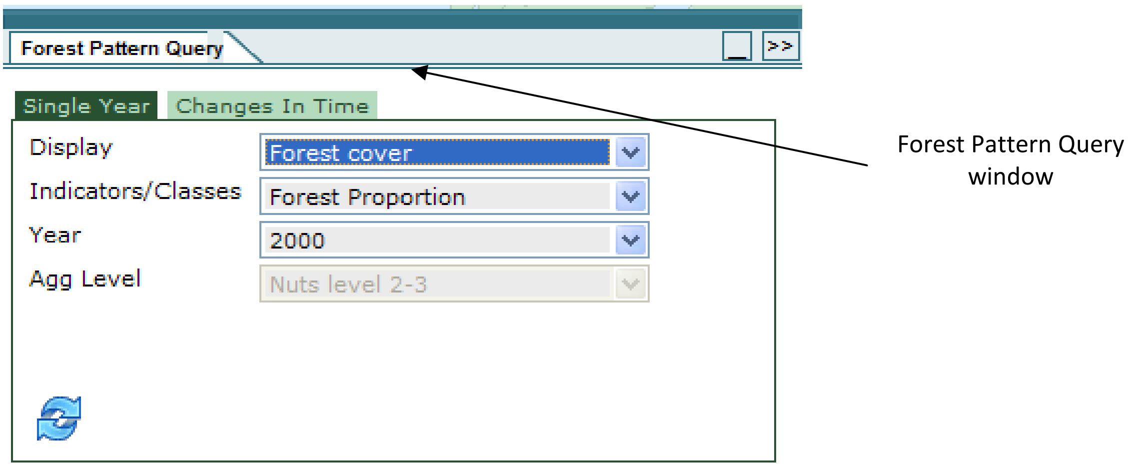


The static raster maps available in the content window are: 1) Forest pattern (CLC 2000); Forest Pattern (CLC 1990); 3) Forest cover change (1990-2000); and 4) Forest / Non Forest Map.

In order to allow user accessing the layer from other data source, the predefined pilot layers from European Soil Data Centre (ESDAC) were made available in the content windows. This was done only for testing and demonstration purposes including only the several most relevant layers from ESDAC for forest user's point of view.

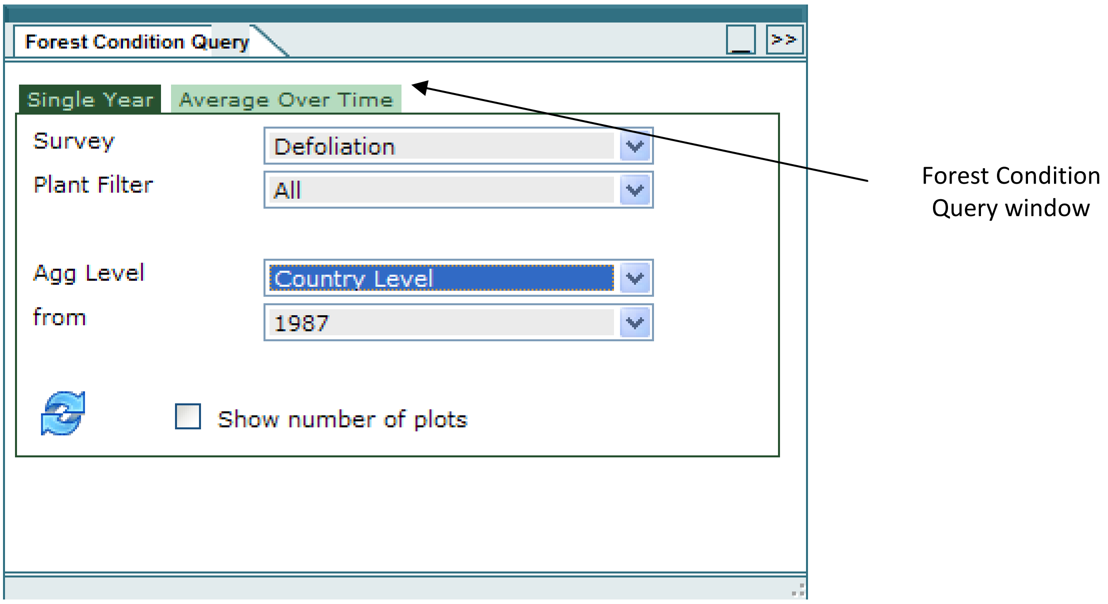
At the bottom of the content window there are 2 expanding query tabs which allow user to select the preferred vector layers based on different criteria and options. The selected layer is displayed on the screen and automatically included in the content window.

Forest Pattern Query. It is designed to query the vector layers related to forest pattern and fragmentation indicators. Two tabs of the query give user the possibility to request the layer for the single year or the layer which displays the change of forest pattern and fragmentation indicator in time (1990-2000). In the query window user can choose the indicator to be displayed (for instance: Forest cover, Forest pattern classes, Core forests, Edge forest, Forest connectivity), the indicator / classes (for instance: for Forest connectivity indicator user can choose the classes – 1 km Dispersal distance, 5 km Dispersal distance, 10 km Dispersal distance, 25 km Dispersal distance), the year (1990 and 2000) and the aggregation level (country, NUTS 1 and NUTS 2-3). For several indicators only the NUTS 2-3 level is available. The reason why NUTS 2 and NUTS 3 levels are not separated is that the size of NUTS 3 level in EU countries is very different which makes more logical for some countries to display the indicator on NUTS 2 while for others – on NUTS 3 level.

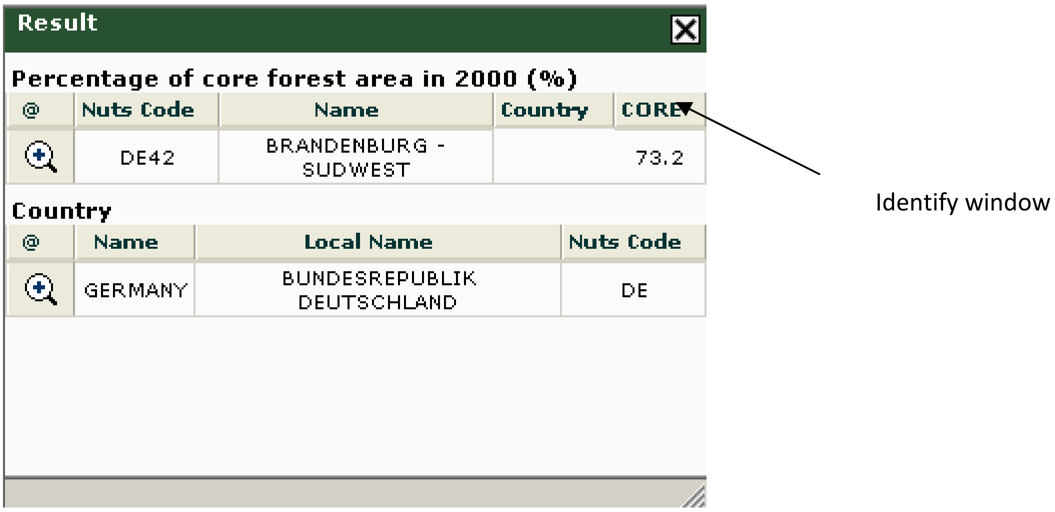


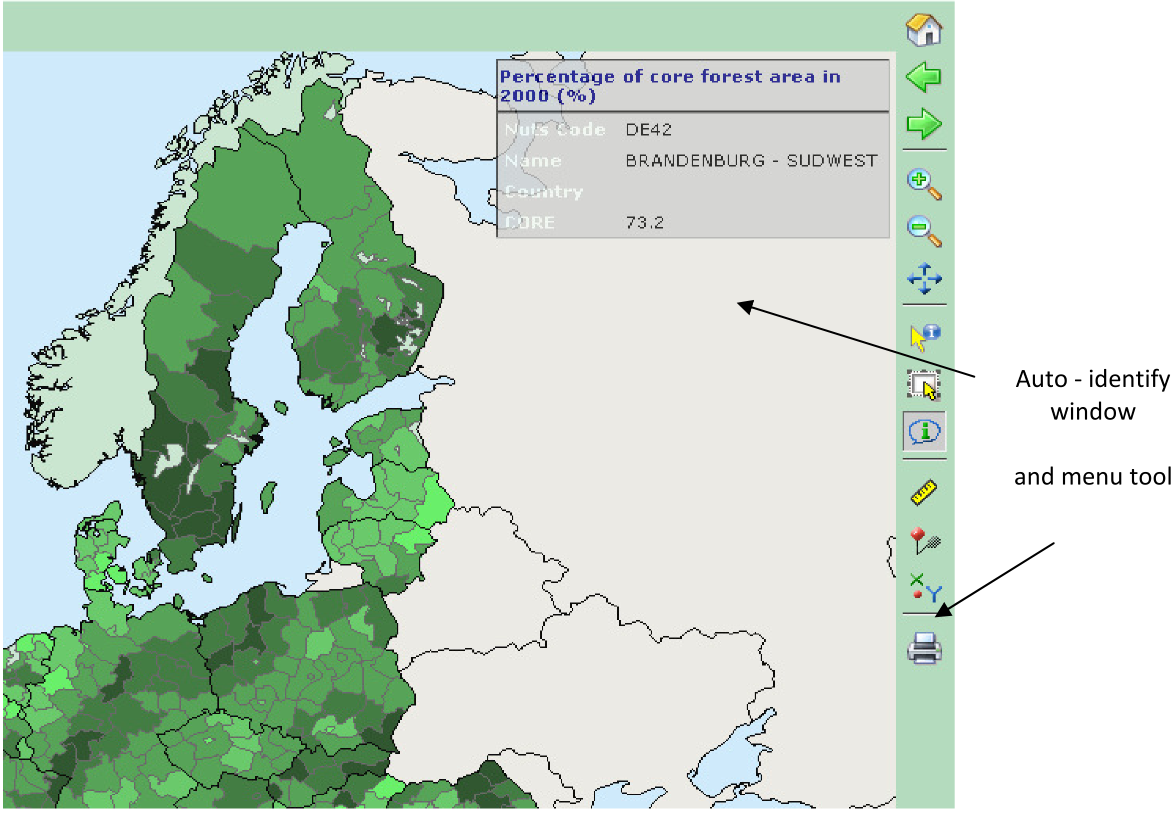
Forest Condition Query. It is designed to query the vector layers related to forest condition. As in Forest Pattern Query two tabs are available to user for selection: 1) Single Year tab; and 2) Average over Time tab. The Single Year tab gives user the possibility to select the forest condition layer for the single year while the Average over Time tab provides the option to select the forest condition layer as the average data from the selected year range.

In the query window user can choose the survey to be displayed (for instance: Defoliation or Discoloration), the Plant Filter (for instance: All, Forest type, Genus, Species), the year (the range is between 1987 and 2005) and the aggregation level (country, NUTS 1 and NUTS 2). In addition the user can select the option to display the number of plots per aggregation unites.



After querying the preferred layer user can navigate it using the menu bar on the right site of the main window. Besides the zooming and selection functions there are "identify" or "auto identify" tools which allow user to see the statistical data of the selected NUTS area. The "auto identify" option automatically displays this information moving through on the screen with the mouse, while "identify" - the same information only on selected NUTS area.





User requirements & Functional- Technical Requirements. The user & functional / technical requirements depends on the design of the EBONE data management framework and on twofold approach as it was mentioned in the software - interface description (in case EBONE will envisage it): 1) the possibility to aggregate WMS/WFS services existing in other systems; and 2) publishing other mapping information that doesn't exist in other mapping systems. In the table bellow the user & functional / technical requirements are presented taking into account the extended possibilities of EBONE data management framework, nevertheless some of them are not implemented in EFDAC Map Viewer but are being considered to implement in the nearest future. Therefore some of the proposed requirements could be skipped or postponed till later – more advanced development phase of EBONE data management framework. However the basic principals / requirements could be the same. Certainly, the mentioned standards could be different as they are suggested taking into account the further EFDAC development, which implies that if, EBONE data management framework considers these requirements, it would be easier in the future for EBONE to communicate with the similar systems, particularly in forest thematic area.

6.4.4 User & Functional – technical requirements

No	Description	Comments
User requirements		
1	Users shall be able to build search criteria in order to query available data.	
2	Users shall be able to view the layer as the result of the query and navigate it on the Map Viewer.	
3	Users shall be able to access a WMS entry point via a Map Viewer interface.	
4	Users shall be able to access a WFS entry point via a Map Viewer interface.	
5	Users shall be able to combine data from different WMS/WFS available services.	Extended possibilities
6	Users shall be able to save the map context for later use.	
7	Users shall be able to create additional layers using a combination of different datasets from different services.	Extended possibilities
8	Users shall be able to download/export a combined dataset with data and save it locally.	Extended possibilities

Functional – technical requirements		
9	The supported browsers shall be Internet Explorer (defined minimum version), Mozilla Firefox (defined minimum version), and Chrome (defined minimum version) on any Operating System that support these browser implementations.	
10	The system shall use as Web Server the Apache HTTP Server.	
11	The application shall be hosted by a Linux Operating System – CentOS 5.0.	
12	The application shall use Java Enterprise and database components namely: Servlets, JSP pages, Web Service implementations, EJB and JDBC, in order to supply all the required functionality.	
13	The system shall use a spatial Database - Postgres/PostGis or Oracle.	
14	The system database could be hosted by a Linux Operating System.	
15	The TCP/IP protocol shall be used for all the communication between EBONE system and external systems.	Extended possibilities
16	The interface between the application and database shall be governed using the JDBC and OCI protocols.	
17	Within the application, all sub-systems (i.e. GIS engine, Interoperability Layer) shall communicate using Java interfaces.	Extended possibilities
18	The application sub-systems Harvester and Publisher shall use the HTTP Get/Post or SOAP/WSDL (web services) protocols for external systems communication. Eventually other (e.g.: FTP) could be considered.	Extended possibilities
19	The system shall use the XML language format for message data & metadata exchange.	
20	The system shall publish geographic datasets and maps through WMS and WFS to other systems.	Extended possibilities
21	The system shall use a Map Server to publish WMS and WFS services within it's system, acting as central point or proxy service.	Extended possibilities
22	The Map Viewer shall comply with the ISO/OGC 191xx/TC211 standard for geographical information exchange/communication.	
23	The Map Viewer shall communicate with relevant and available systems related to biodiversity.	Extended possibilities
24	The Map Viewer shall use the most recent versions OGC standards for publishing geographical data, named as WMS, WFS and WCS.	Extended possibilities
25	The Map Viewer shall use the open source software GeoServer as a map server (UMN MapServer could be as alternative).	
26	The Map Viewer shall be an independent component with interfaces available for integration.	Extended possibilities
27	The Map Viewer shall combine several data layers from different information sources.	Extended possibilities
28	The Map Viewer shall allow export and save result data combination.	Extended possibilities
29	The Map Viewer shall have a public internet address.	
30	The Map Viewer shall have a public mapping interface address.	
31	The Map Viewer shall use the Internet for all communications with other systems.	

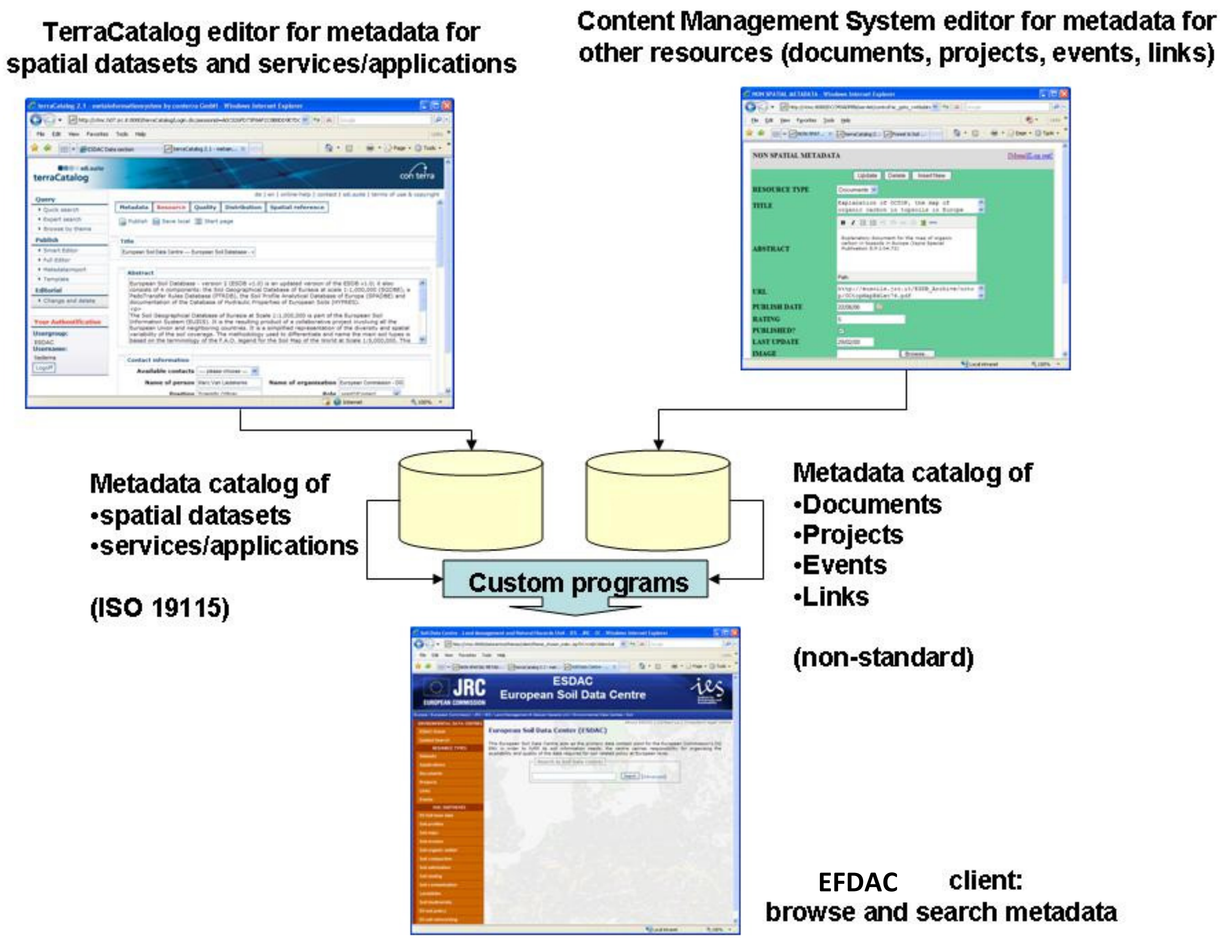
Access of EFDAC Map Viewer application / service through Metadata catalogue

The EFDAC technical platform allows the insertion and publication of meta-data in a JRC-central catalogue for a number of different resource types (datasets, services, applications, documents, reports, projects, and links). Such a platform currently consists of a lightweight system that allows users through the web to browse and search for forest resources from a local catalogue of metadata. The metadata catalogue is being populated with metadata of JRC in-house data and is built around a metadata model that, for geo-spatial data, is compliant with OGC ISO CSW 2.0.2 specifications.

It is expected in the future to shift the EFDAC Catalogue to GeoNetwork as the open – source platform which would increase the possibilities and accessibility of the current EFDAC metadata catalogue. Therefore the brief description of the current EFDAC metadata platform is provided to show the access of EFDAC Map Viewer application / service through the Metadata catalogue.

The EFDAC metadata catalogue is queried by the lightweight client directly through Java Servlets to the database. TerraCatalog gives the possibility to expose the geographical data through an OGC CSW Service version 2.0.2 which at this stage isn't exposed externally. While non geographical data is not interoperable for externals systems. The EFDAC metadata client interface gives the user the possibility to search metadata by forest subthemes or resources types. Through the metadata record the user can directly access the data source. Metadata for datasets and services are inserted in a commercial meta-data catalogue (ConTerra) while metadata for documents (reports), projects, events and links are inserted in a JRC developed custom catalogue adapted to specific requirements. The technical platform reads meta-data from these catalogues in order to provide browse and search capabilities into the inserted metadata.

Metadata for datasets and services are inserted in a commercial metadata catalogue (ConTerra) that is compliant with ISO19115 and ISO19119, recommended by INSPIRE. The system uses two different repositories: one for the geographical metadata stored by TerraCatalog software and another one for the non geographical information in two different table spaces within the same Oracle database. Bellow is presented the schema of EFDAC metadata platform.



The metadata record in the client interface shows the concise information about the data source and provides the link to the application which handles the data, link to download service and link to detailed information about the data (see the example bellow).

European Forest Spatial Pattern 2000

Title	European Forest Spatial Pattern 2000
Abstract	The pattern map is available over Europe for year 2000 with a spatial resolution of 100 m. It contains 7 forest pattern classes: 1) "core forest", 2) "edge" of core, 3) "perforation edge" (inside core patches), 4) "branch", 5) "connection: loop", 6) "connection: bridge" (between core patches), and 7) "islet" (small forest patch). The map has been obtained from the binary forest cover map of CORINE Land Cover 2000. The method is based on morphological image filtering techniques. On the basis of pattern maps, Forest Pattern and Fragmentation indicators were developed.
Keywords	forest spatial pattern; fragmentation; connectivity; Forest biodiversity
Geographic Extent	**West:** -15.0 **East:** 40.0 **South:** 35.0 **North:** 75.0
Date	**Publication:** 2009-02-11 **Revision:** 2008-12-07 **Creation:** 2007-05-22
Online Resource	**Information:** http://forest.jrc.ec.europa.eu/forest-pattern/pattern-products **Download:** http://forest.jrc.ec.europa.eu/download/data/MSPA-Maps **Application:** http://efdac.jrc.ec.europa.eu/viewer
Responsible Party	**Organisation Name:** European Commission, Joint Research Centre, Institute for Environment and Sustainability
Constraints	**Access constraints:** copyright
Topic Category	environment
Language	eng

Use Cases. Three use cases according to different queries and results are described bellow. Nevertheless they are characteristics to EFDAC data; however the same principals and behaviour could be set up for EBONE as well. The first use case describes the raster map query while the second one – the forest pattern and fragmentation indicators query. The last use case shows the forest condition data query with the option to show the number of forest plots per NUTS area.

Use Case: Raster Maps Analysis	
Step	**Use Case Description**
1	User accesses the EFDAC Metadata Catalogue and searches for "Forest Spatial Pattern" (http://efdac-catalog.jrc.ec.europa.eu). From the search results user selects the metadata record "European Spatial pattern 2000". *(Note: in EBONE the search could be performed trough catalogue interface / or other entry point*).
2	From the metadata record "European Spatial pattern 2000" user accesses the download page (http://forest.jrc.ec.europa.eu/download/data/MSPA-Maps) of European Spatial Pattern 2000 (geotiff format) and later accesses the EFDAC Map Viewer application (http://efdac.jrc.ec.europa.eu/viewer/) which handles the spatial pattern and other maps stored in JRC. The application shows the Forest Spatial Pattern 2000 raster map.
3	Using the search tool user searches the area of interest (for instance: NUTS 3 level area of GIRONDE in France) and views the forest spatial pattern in this are. In order to see the pattern in more details user zooms the area till the appropriate level using the zoom tool of the menu bar. User can see the forest spatial pattern classes (Core Forest, Islet Forest, Edge Forest, Branch Forest, Perforation edge Forest, Connection Forest: Loop, Connection Forest: Bridge).
4	In order to see the forest cover change in the same area of interest during 1990 – 2000 user selects from the content windows the Forest Cover change 1990-2000 raster layer, which provides the spatial distribution of the gross forest loss (area out of forest), of the gross forest gain (area into forest) and of stable forest in the period 1990-2000. Distinctions between net and gross changes in forest area are important in the biodiversity context. The user deselects the Forest Cover change 1990-2000 raster layer and sees the previous forest spatial pattern layer 2000 analyzing the same area of interest after change took place.
5	In order to see the forest layer on the same NUTS area, user deselects the current layer and from the content window selects the Forest / Non – Forest Map 2000.
6	Selecting and deselecting Forest Spatial Pattern 2000, Forest Spatial Pattern 1990, Forest Cover change 1990-2000 and Forest / Non – Forest Map 2000 raster maps user analyses the are of interest in terms of forest fragmentation, forest gain / loss and etc.

Use Case: Forest pattern and fragmentation indicators	
Step	**Use Case Description**
1	User accesses the EFDAC Map Viewer application http://efdac.jrc.ec.europa.eu/viewer and expands "Forest Pattern Query" window.
2	In the "Single Year" tab of the "Forest Pattern Query" window user enters the search criteria, for instance: Display: "Forest Pattern Classes"; Indicator / Classes: "Core forest proportion"; Year: "2000"; Aggregation level: "Nuts level 2-3" and runs the query. The "Percentage of core forest area in 2000 (%)" vector map appears on the screen.
3	In order to know information about the core forest user selects the info icon besides the title of the layer and opens the "layer information" window with the description of detailed information about the map.
4	User selects the "auto identify" bottom from the menu bar and views the name as well as the percentage of core forest area in 2000 of the selected NUTS area.
5	In order to see change of core forest during the time period 1990-2000 the user enters the search criteria in the "Changes in Time" tab of the "Forest Pattern Query", namely: Display: "Change of core forest"; Indicator / Classes: "Loss of core area" and runs the query. The "Gross core forest loss (PCLOS) 1990-2000 (%)" vector layer appears on the screen.
6	Selecting and deselecting the "Percentage of core forest area in 2000 (%)" layer and "Gross core forest loss (PCLOS) 1990-2000 (%)" user analyses the area of interest in terms of core forest area.

Use Case: Forest condition	
Step	**Use Case Description**
1	User accesses the EFDAC Map Viewer application http://efdac.jrc.ec.europa.eu/viewer and expands "Forest Condition Query" window.
2	In the "Single Year" tab of the "Forest Condition Query" user enters the search criteria, for instance: Survey: "Defoliation"; Plant Filter: "All"; Year: "2005"; Aggregation level: "Nuts level 2", checks the option "Show number of plots" and runs the query. The "Crown defoliation (2005) survey plot averages for all sampled trees" with the number of plots in each NUTS 2 area appears on the screen.
3	In order to know information about the defoliation layer user selects the info icon besides the title of the layer and opens the "layer information" window with the description of detailed information.
5	In the "Average over Time" tab of the "Forest Condition Query" user enters the search criteria, for instance: Survey: "Defoliation"; Plant Filter: "All"; Year: "from 2000 till 2005"; Aggregation level: "Nuts level 2", checks the option "Show number of plots" and runs the query. The "Crown defoliation (2000-2005) survey plot averages for all sampled trees" with the number of plots in each NUTS 2 area appears on the screen.

6	In order to see what is the soil erodibility condition in the area of interest (for instance, where the defoliation is high) user selects the "Soil erodibility" layer from the "Access to soil data (ESDAC) title in the content window. The "Soil erodibility" layer appears on the screen. Selecting and deselecting other available layers in the content windows user can quickly see and analyze the area of interest in terms of forest and soil condition.

6.4.5 Implementation

The EFDAC portal with its web GIS viewing possibilities will be used within EBONE as test portal for the biodiversity data. The know-how of existing projects and European efforts will be used to establish an exemplary web portal for biodiversity information based on landscape sample based habitat and species monitoring.

The steps to be taken in this approach:

- Establishment of WMS services for the field data.
- Establishment of WMS services for the resulting biodiversity indicators based on the European Environmental Zones and/or Strata.
- Testing of the web portal.

6.5 Helpdesk

The introduction to the mobile field mapping application will be given at the field training course in Spain and Romania.

A helpdesk will be needed to focus the communication between the users and the programming team. A first level support will be created by Umweltbundesamt and Alterra – EBONE@umweltbundesamt.at – to collect the problems. Problems with the use of the application should be solved at this level. Technical problems will be directed to the programmers.

At the internal project web site at http://www.surfgroepen.nl/ a wiki section for posting questions and providing help will be generated.

6.6 Cost aspects

At the current phase of the project only free and open source components are used for the EBONE data management framework. For the field database license costs for Microsoft Access have to be taken into account.

A detailed cost estimate will be made by the updated report.

6.7 Time plan

The following activities described in the sections above are planned to be implemented in 2010. 2010 will be the test phase for the EBONE data management framework.

Tasks 2010	1	2	3	4	5	6	7	8	9	10	11	12
Technical specification	■	■	■								■	■
Common domain data model			■	■	■	■						
Field Computer	■	■	■	■								
Field Database	■	■	■	■								
WFS (for selected data sources)				■	■	■	■					

WMS (for selected data sources)				■	■	■	■					
CSW (e.g. geonetworks)				■	■	■	■					
Data warehouse						■	■	■	■			
Data portal (e.g. EFDAC)									■	■	■	
Evaluation of framework											■	■

In 2011 adoptions and improvements of the EBONE data management framework will be made based on the assessment and evaluation of the test phase in 2010.

7 First implemented steps

Based on existing habitat mapping data within the EBONE consortium the steps of the data flow model were implemented using a simple Access database. Example data from existing habitat mapping projects - the British Country Side Survey (UK), the North Ireland Country Side Survey (N-IRL), NILS (Sweden), SINUS (Austria) – as well as from landscape squares mapped according to the EBONE habitat protocol – example data from France and Israel – were used to do the test. The spatial distribution of the test data for the data integration test is shown in the map below.

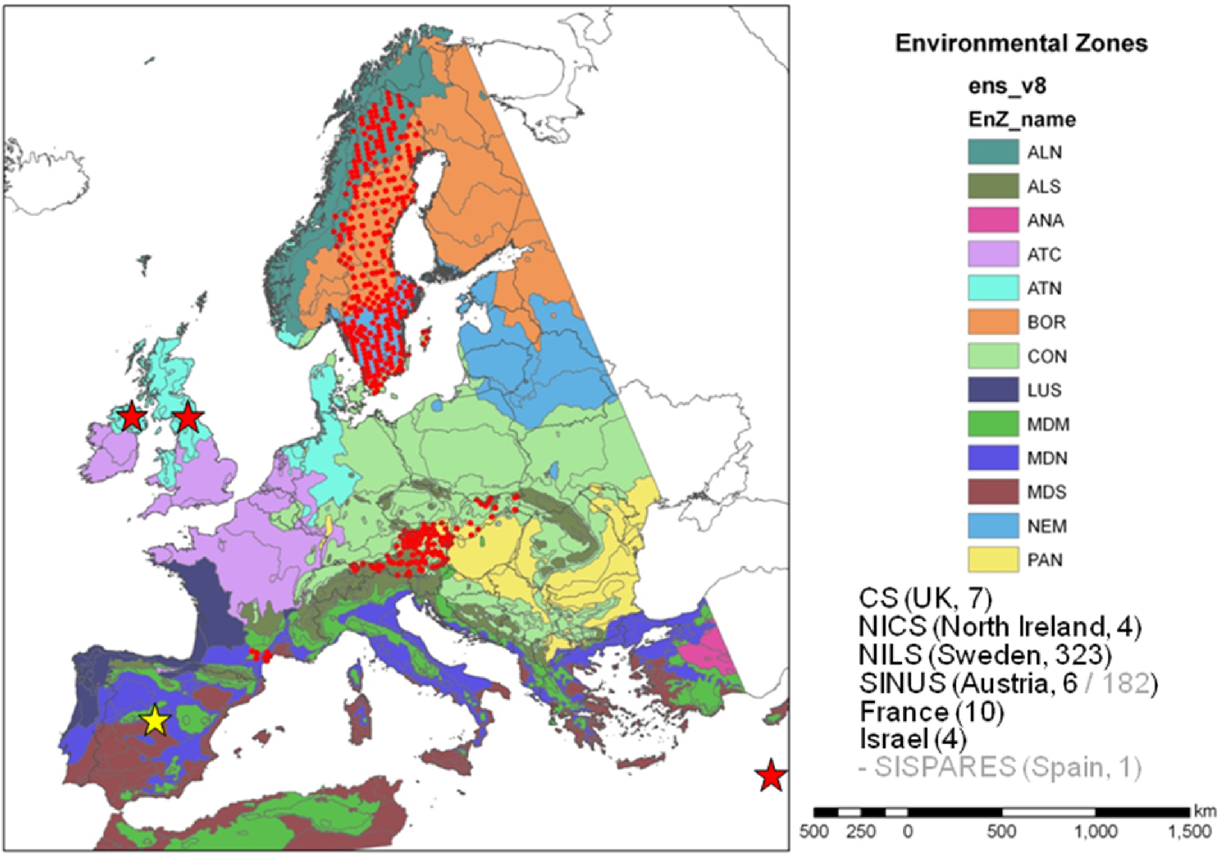


Fig. 21 Distribution of location of test data for integration

Based on the example data a simple common domain model was discussed consisting of the following schema elements: a) landscape element (or polygon) having a name (identifier), area (in ha), GHC (resulting from the transformation), and optional the original habitat recording and spatial information, b) the landscape square (or plot) having a name (identifier), total area (in ha), assignment to environmental strata and zone, and optional spatial information about the location (often issue of data policy), and c) Environmental Stratum (EnS) having a name (identifier), total area of the stratum (in square kilometre) and the spatial information about the extent. This basic common domain model is the starting point for the further work on it within the work package which will be carried out during spring 2010.

Simple calculations of the average proportion of GHCs on the level of the EnS were performed and an example for the strata Alpine North 1 and 2 is given in Fig. 22. Mean values and standard deviation was calculated using SQL (not shown in the diagram).

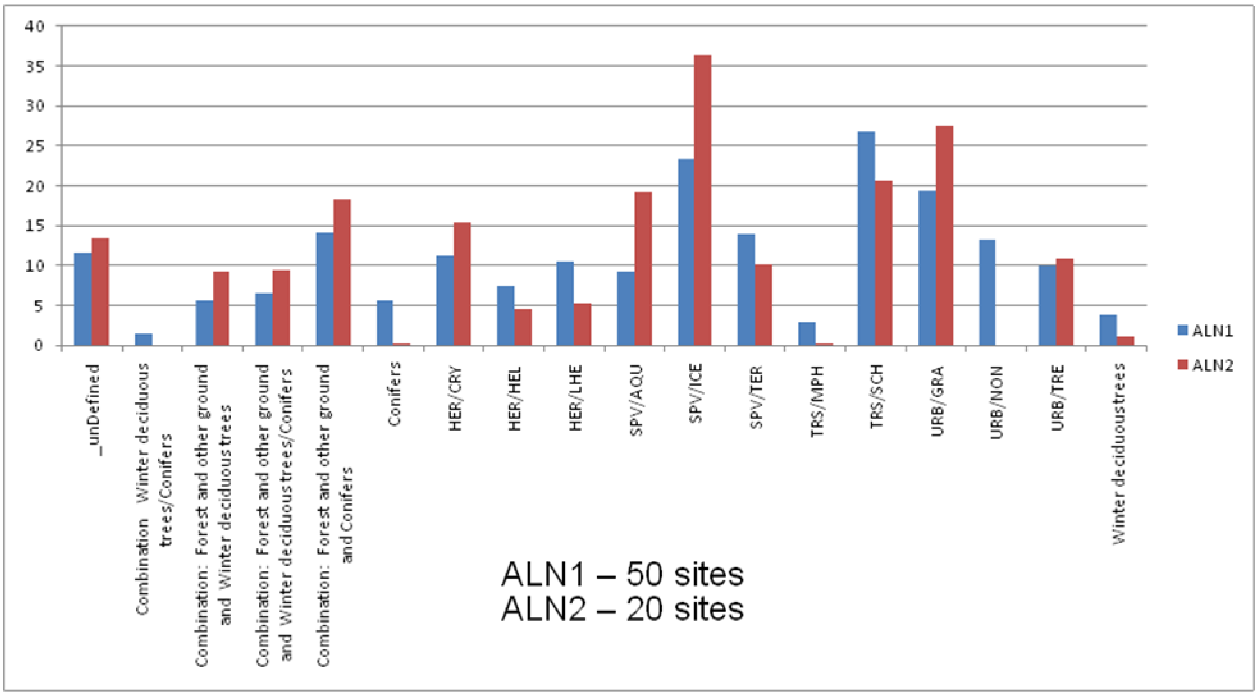


Fig. 22 Exemplary calculation of the mean share of GHC per EnS on the basis of the harmonised example data for the two strata Alpine North 1 (ALN1) and 2 (ALN2). The number of landscape squares for the EnS is given in the diagram as the number of sites.

8 Test cases existing data

In the following section none EBONE data sources are described which will be used as test data sources for the EBONE data management system.

8.1 NILS

NILS monitors the landscape with random sampling in 631 permanent sample plots, systematically distributed in Sweden. One fifth of the plots is inventoried each year, which means every plot is sampled in a 5-year interval.

NILS uses landscape squares (5x5 km) for general descriptions with a central kilometre square (1x1 km) for more comprehensive measurements.

Strata divisions

Sweden has been divided into ten geographical strata differing in sampling density. The divisions provide the basis for concentrated random sampling in certain areas, e.g. cultivated land and the alpine region.

Aerial Photo Interpretation

Comprehensive and detailed interpretations are done in the kilometre square (1x1 km). A method of how to interpret the landscape square (5x5 km) is under development.

Field Inventory

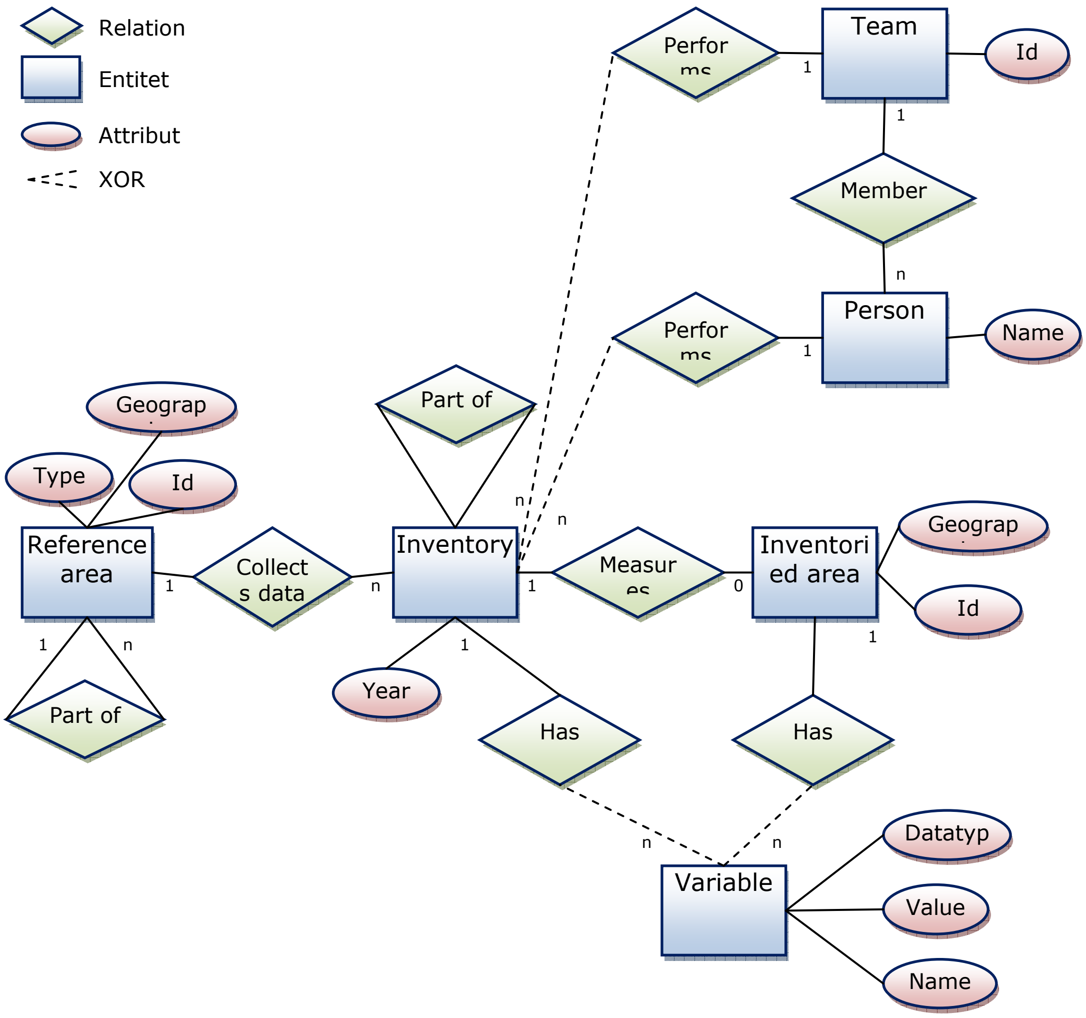
The field inventory adds information that cannot be obtained from the aerial photos - among other things species level information. Both comprehensive inventory of the ground cover and land use, as well as detailed description of the vegetation is performed in systematically placed sample plots within the kilometre square.

Line Inventory

Line inventory provides good estimates of length and quality of linear elements in the landscape (such as roads, hedges and ditches).

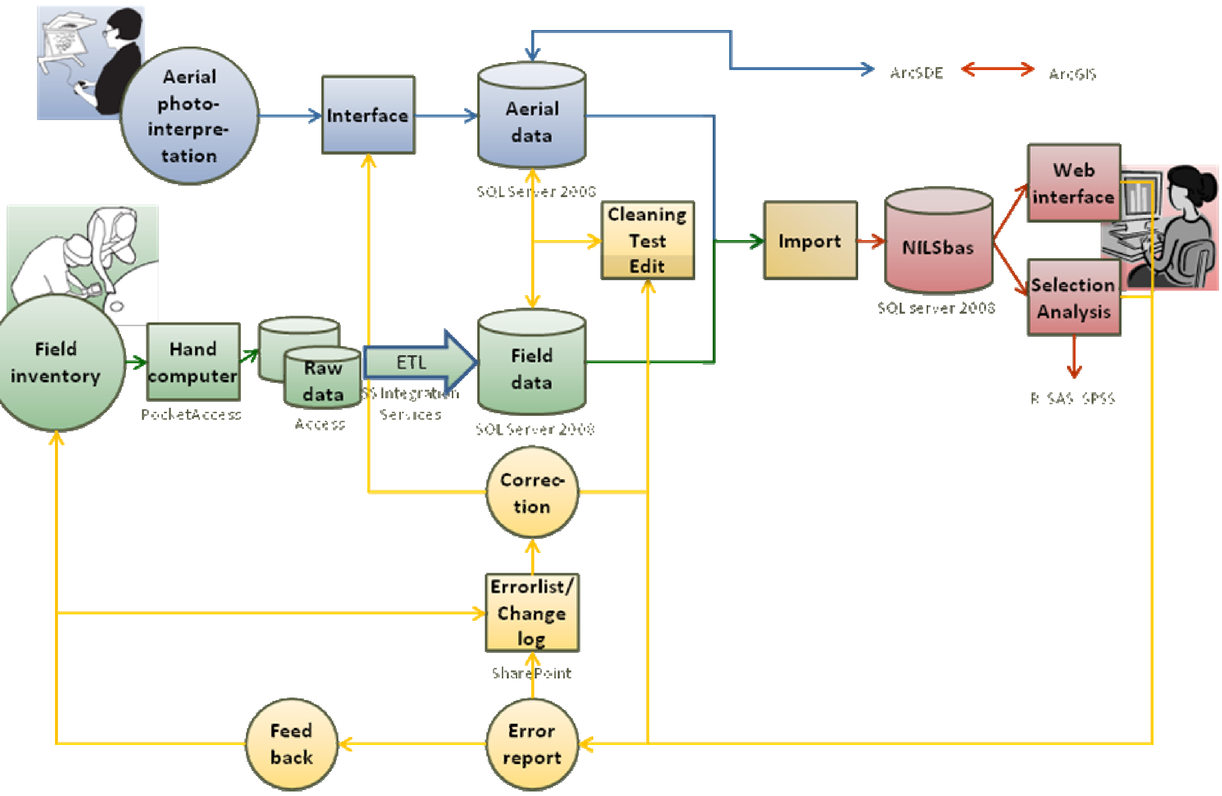
8.1.1 Data structure

The data model is still under construction.



8.1.2 Data management

Conceptual chart of the system architecture in NILS. Parts that are not implemented at present are shaded in grey. Two separate databases are used for editing collected data from aerial photo interpretation (blue) and field inventory (green), whereas NILSbas will be the base for data selection and analysis (red). Data quality is assured in several steps (yellow).



8.1.3 Data transformation

The transformation is described in the file GHC_NILS Conversion_Allard_ 15_April_2009.xlsx.

8.1.4 Data policy issues

An official data policy description is under development. It will be discussed by the board 23 mars and I can send it after that.

8.2 SINUS

The project 'Spatial Indices for Land Use Sustainability (SINUS)' was designed to develop reliable, operational, and spatially explicit indicators of practical use in long-term monitoring and assessment of ecological sustainability of Austrian cultural landscapes. The use of landscape structure attributes for this assessment was tested and implemented in an assessment procedure. The main research question was: which attributes of the land-use mosaic have to be recorded to describe the landscape with respect to its ecological sustainability?

Land use and landscape structure

From the variety of the Austrian Cultural Landscapes 120 1 x 1 km² sample plots were chosen using a stratified random sampling procedure. Base unit for the mapping was the landscape element as well as the landscape. In a field survey for the whole landscape plot every landscape element was delineated on a base map. Attributes of landscape structure, the hemerobiotic state, the trophic level, plant species richness, origin type and dynamic were recorded for every single landscape element. The spatial data were stored in a GIS that was linked with the landscape structure database JOKL.

Land cover and landscape structure

Different methods of satellite imagery were tested (e.g. sub-pixel analysis, watershed segmentation, etc.) to select the most efficient for this huge data set. The methods were applied to the area cover satellite images of Austria. Landsat TM-5 images were used. The combination of an innovative segmentation method (region-growing algorithm) and classification procedure (knowledge based classification by using additional attributes like shape and spatial distribution of the segments) resulted in an efficient use of the resources. The method was implemented in a stand-alone application under a UNIX platform.
As reference data the fine scaled landscape structure mapping (scale 1 : 10.000) as well as the information about the landscape types of the Austrian Cultural Landscapes Classification were used (scale 1 : 200.000; Wrbka et al. 2002).

Indicators

The data analysis was performed in a combination of a bottom-up and a top-down approach. The landscape was quantified by using landscape ecological relevant landscape metrics (see Fragstats; McGarigal & Marks 1994). On the basis of a comprehensive literature review of 'sustainability indicators on the landscape level' as well as the interpretation of the results of the classification of the landscape types and the respective description (fact-sheets), an assessment of the structural configuration of the landscape types was performed. Land use mosaic, fragmentation and connectivity, as well as the landscape characteristics were the main criteria in this step. The resulting indicators were also used for the sustainability assessment of the specific landscape

The degree of anthropogenic influence on a landscape, as measured by the concept of the hemerobiotic state, is an aggregated indicator on the landscape level. The relation between the attributes describing landscape structure - indicators that can easily be measured - and hemerobiotic state - a very complex and aggregated indicator - was analysed. On the basis of these results a system to assess the sustainability of certain landscapes was developed. Base data for the analysis were the results of the fine-scale landscape structure mapping. The analysis of the landscape structure (= 'landscape metrics') was done by using the Patch Analyst 2.2 (Elkie et al. 1999) under ArcView 3.3. The relation between the hemerobiotic state and the variables describing landscape structure were tested on the landscape element level as well as on the landscape level.

Identifying sustainable Austrian Cultural landscapes

Statistical approach - RegSust

Because detailed Austrian wide data about the degree of anthropogenic influence on the landscape level are missing or are only present for certain ecosystems, (see Grabherr et al. 1998) the hemerobiotic state of the Austrian Cultural Landscapes was predicted on a 1 x 1 km grid base using ordinal regression techniques and using the sample plots as sample for the analyses (n=132). The predicted hemerobiotic state was the input for the statistical sustainability assessment. Predictors used included variables on the bio-geographic characteristics, variables describing the landscape pattern and configuration, variables on the characteristics of the segments, and variables describing the fragmentation and connection of the landscape in supra regional networks. The variables on the landscape structure and land use were derived from the land cover map, which was produced within this research project.
The resulting model ($R^2=0.535$, $p<0.001$) was applied to the whole data set and the results were visualised. Based on the modelled hemerobiotic state the sustainability assessment was performed for the cultural landscape type groups, as well as for the cultural landscape type series. The deviation of a certain landscape cell from the average hemerobiotic state of the respective spatial reference unit was determined and classified. The result, the sustainability indicator RegSust, was classified into five classes ranging from -2 (strong negative deviation, reddish) to +2 (strong positive deviation, greenish). Landscape cells showing no deviation from the average hemerobiotic state (0) were displayed as light yellow areas.

Rule based approach – FuzSust

On the basis of a comprehensive literature review and expert knowledge, a set of rules for every cultural landscape type group was defined, describing sustainable land use. Sustainability was described by using the concept of linguistic variables, which allow the use of the terms 'more' or 'less'. Thus different shades of a term could be expressed. The rules were transformed into membership function (μ) using logistic and triangular functions. By relating these rules (approximate reasoning) the membership value for the linguistic variable 'highly sustainable' was determined and visualised. Basis data for the formulation of rules were attributes describing the landscape structure and configuration which were derived from the Austrian land cover map, as well as additional data which were derived from spatial relevant data sets (e.g. digital elevation model, road and stream networks, etc.). All variables were calculated for a grid of 1 x 1 km landscape cells. Each landscape cell was assigned to the dominant landscape type group and landscape type series of the Austrian cultural landscape classification.

8.2.1 Data structure

Data processing and management

The processing and management of the base data, as well as the documentation and visualisation of the results were major issues of this research project. The usefulness of data and results for future projects is largely depending on that.

The presentation of the data and the result is done on four different levels:

(a) visualisation of the base data using the map tool developed under ARC/Info

(b) analogue and digital preparation of the results as maps (pdf-files and gif-images)

(c) presentation of the project results on the project homepage

(d) presentation of the project results on a CD (final report and maps).

8.2.2 Data management

The data management is done using a Microsoft Access application. The JOKL application follows the structure of the SINUS field manual. The spatial data are stored as ESRI shape files identified by a unique identifier.

The application provides interfaces for data entry, quality control, and data query.

8.2.3 Data transformation

For the following land use types transformation into GHCs was defined:

DEF_Nutztyp_GHC				
ID	**NutztypKlar**	**Land use typ**	**C-Code**	**GHC**
1	Acker Hackfrucht extensiv	root crop extensive	AHE	CUL/CRO
2	Acker Hackfrucht intensiv	root crop intensive	AHI	CUL/CRO
3	Acker Hackfrucht mäßig int.	root crop medium intensive	AHM	CUL/CRO
4	Getreideacker extensiv	corn fields extensive	AE	CUL/CRO
5	Getreideacker intensiv	corn fields intensive	AI	CUL/CRO
6	Getreideacker mäßig intensiv	corn fields medium intensive	AMI	CUL/CRO
7	Weingarten intensiv	vineyard intensive	WGI	CUL/WOC
8	Weingarten mäßig intensiv	vineyard medium intensive	WGM	CUL/WOC
9	Wald Forst alt	timber plantation old	WFA	FPH
10	Wald Forst jung	timber plantation young	WFJ	FPH
11	Allee alt	avenue with old trees	ALLA	FPH/DEC
12	Allee jung	avenue with young trees	ALLJ	FPH/DEC
13	Baumweiden alt	pasture with old trees	BWEA	FPH/DEC
14	Baumweiden jung	pasture with joung trees	BWEJ	FPH/DEC

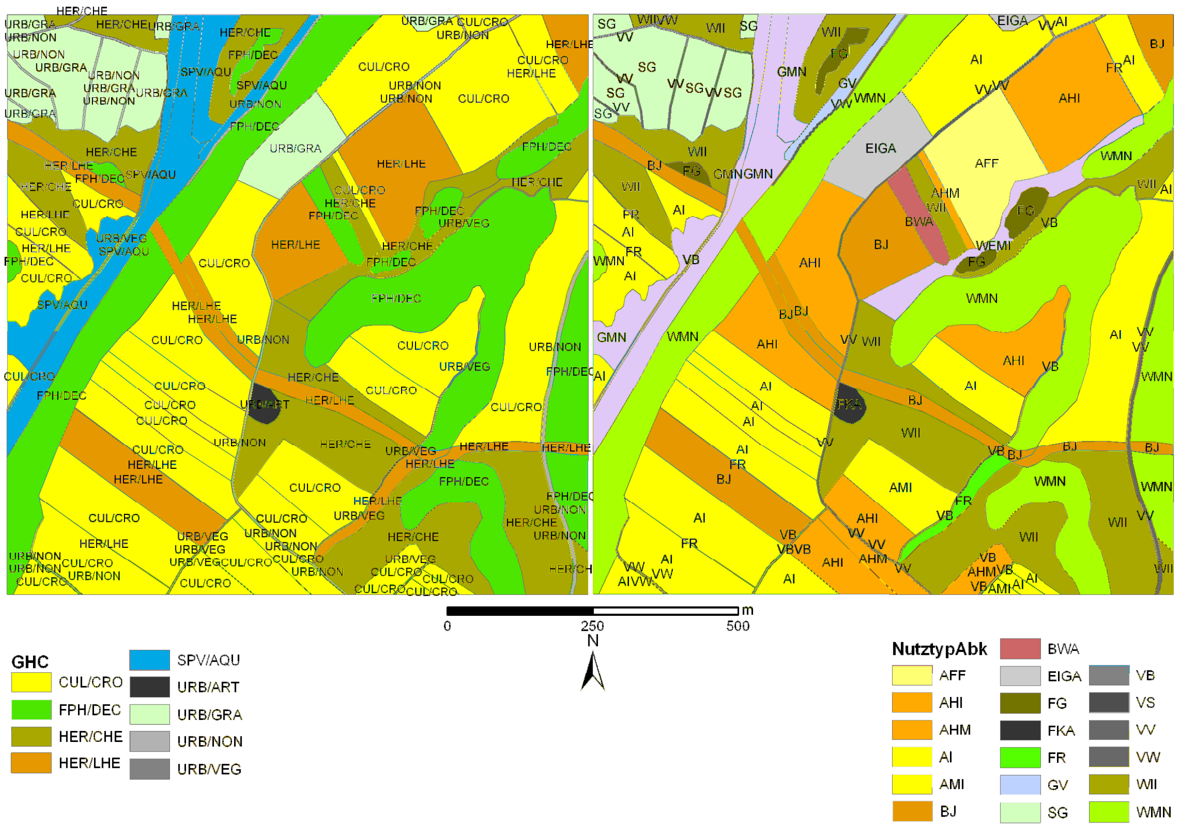
DEF_Nutztyp_GHC				
ID	**NutztypKlar**	**Land use typ**	**C-Code**	**GHC**
15	Baumwiese alt	orchard old	BWA	CUL/WOC
16	Baumwiese jung	orchard young	BWJ	CUL/WOC
17	Einzelbaum alt	old solitary tree	EBA	FPH/DEC
18	Einzelbaum jung	young solitary tree	EBJ	FPH/DEC
19	Feldgehölz	small woodlot	FG	FPH/DEC
20	Hecke Baum	hedgerow of trees	HB	FPH/DEC
21	Hecke Strauch	hedgerow of shrubs	HS	FPH/DEC
22	Wald mäßig naturnah	seminatural forest	WMN	FPH
23	Wald naturnah	natural forest	WN	FPH
24	Sonderbiotope natürlich	special biotop - natural	SONN	HER/CHE
25	Weide extensiv	pasture extensive	WEE	HER/CHE
26	Weide intensiv	pasture intensive	WEI	HER/CHE
27	Weide mäßig intensiv	pasture medium intensive	WEMI	HER/CHE
28	Wiese extensiv	meadow extensive	WIE	HER/CHE
29	Wiese intensiv	meadow intensive	WII	HER/CHE
30	Wiese mäßig intensiv	meadow medium intensive	WMI	HER/CHE
31	Acker mit Feldfutteranbau	forage crops	AFF	CUL/CRO
32	Brache jung	young fallow land	BJ	HER/LHE
33	Brache mit Staudenflur	old fallow land with tall herbs	BS	HER/LHE
34	Feldraine	field margin	FR	HER/LHE
35	Fließgewässer mäßig naturnah	stream seminatural	GMN	SPV/AQU
36	Fließgewässer naturnah	stream natural	GN	SPV/AQU
37	Fließgewässer verbaut	stream artificial	GV	SPV/AQU
38	Stillgewässer natürlich	lake natural	STL	SPV/AQU
39	Stillgewässer naturnah	lake seminatural	STN	SPV/AQU
40	Brache mit Gehölzflur	old fallow land with shrubs	BG	TPH/DEC
41	Blockrand- bzw. Zeilenverbauung verdichtet	blocks	BZV	URB/ART
42	Dorfkern aufgelockert	village vegetated	DFKA	URB/ART
43	Einzelgehöfte und Kleinweiler verdichtet	detached houses paved	EIGV	URB/ART
44	Einzelhausbebauung	one-family h.	EIH	URB/ART
45	flächige Kleinarchitektur	built up element	FKA	URB/ART
46	Industrie- und Gewerbegebiet aufgelockert	industrial sites veg.	IGA	URB/ART
47	Industrie- und Gewerbegebiet verdichtet	industrial sites paved	IGV	URB/ART
48	Lineare Kleinarchitektur	built up element linear	LKA	URB/ART
49	Materialdeponieen	Deposition, land fill	DEP	URB/ART
50	punktförm.Kleinarchitektur	built up element punctif.	PKA	URB/ART
51	Siedlung versiegelt	settlements paved	SV	URB/ART
52	Sonderbiotope künstlich	special biotop - artificial	SONK	URB/ART
53	versiegelte Sonderflächen	other paved areas	VS	URB/ART
54	wassergebundene Sonderflächen	Other unpaved areas	WS	URB/ART
55	Dorfrand	suburb	DFR	URB/GRA
56	Dorfrand aufgelockert	suburb vegetated	DFRA	URB/GRA
57	durchgrünte Einzelhausbebauung	one-family houses veg.	EIHA	URB/GRA
58	Einzelgehöfte und Kleinweiler	detached h.	EIG	URB/GRA
59	Einzelgehöfte und Kleinweiler aufgelockert	detached houses veg.	EIGA	URB/GRA
60	Siedlung grün	settlements, vegetated	SG	URB/GRA
61	periodisches Fließgewässer künstlich	periodic stream artificial	PFK	URB/NON
62	periodisches Fließgewässer natürlich	periodic stream natural	PFN	SPV/AQU

DEF_Nutztyp_GHC				
ID	**NutztypKlar**	**Land use typ**	**C-Code**	**GHC**
63	Stillgewässer künstlich	lake artificial	STK	URB/NON
64	Verkehrsweg versiegelt	paved roads	VV	URB/NON
65	Verkehrsweg wassergebunden	dirt roads	VW	URB/NON
66	Parks und Gärten	Gardens, parks	PG	URB/VEG
67	Verkehrswege begrünt	roads vegetated	VB	URB/VEG
68	Blockrand- bzw. Zeilenverbauung aufgelockert		BZA	URB/ART
69	Dorfkern		DFK	URB/ART
70	Dorfrand verdichtet		DFRV	URB/ART
71	Dorfkern verdichtet		DFKV	URB/ART
72	verdichtete Einzelhausbebauung		EIHV	URB/ART
73	Gehölzplantagen		GP	CUL/WOC
74	Materialentnahmestellen		MAT	URB/VEG
75	periodisches Stillgew. künstlich		PSK	URB/NON
76	periodisches Stillgew. natürlich		PSN	SPV/AQU
77	nicht beschriebener Wald		W	FPH
78	Weingarten extensiv		WGE	CUL/WOC

8.2.4 Data policy issues

Data is available by contacting Dr. Thomas Wrbka (thomas.wrbka@univie.ac.at). A service fee is charged for preparing and handling the requested data from the SINUS-project for projects where the research group is not involved in the consortium.

8.2.5 Data example



8.3 SISPARES - The Spanish Rural Landscape Monitoring System

SISPARES is a monitoring system to study the ecological values and dynamics of the Spanish Rural Landscapes including their characterisation, classification, and recent past and future dynamics (Elena-Rosselló et al., 2005). The specific objectives are to identify and monitor landscape structural elements and processes. The main action was the establishment of a representative Spanish Rural Landscape Network (REDPARES, Bolaños et al., 2003*)* that has 215 plots of 4x4 km based on a stratified simple random sampling. The intensity of sampling, taking into account sample size and the area of the national territory, is around 1/146.

All plots were surveyed through aerial photographs three times: 1956, 1984 and 1998 datasets. The last date photo-interpreted was validated by means of field visits. A new survey is updating the 215 plots using 2007 photograph datasets and new field visits will be completed.

The sampling design was based on *CLATERES* land classification (Elena-Rosselló et al., 1997) that defined 215 biogeoclimatic classes in Iberian Peninsula and Balearic Islands. This classification used a statistical clustering routine identical to ITE land classification, on the basis of physiographic and geographic information, lithology and climate (Climatic data had different source that EnS). CLATERES was constructed by TWINSPAN multivariate analysis (Hill, 1979). Its design was based on a downwards nested classification analysis that assembles a bifocal dendrogramic pattern and enables the classes to be used as a basis for a stratified field sampling (Barr *et al*., 1993). This pattern also allows the upward grouping of neighbouring land classes into higher level categories for data analysis purposes (Smith, 1982). Any new class in the emerging level is the most significant union between the land classes existing in the previous lower level. The methods that made up CLATERES were structure into two phases. In a first phase, the whole Spanish Peninsula and Balearic Islands were classified on a scale of 25x25 km in 13 classes and in a second phase with a resolution of 2x2 km in 215 classes.

8.3.1 Data structure

Each plot of 4x4 km represents the rural landscape of one geoclimatic class and was analysed by delimiting patches of land cover and linear elements of road network from aerial photo interpretation. The scale of photos of 1956 and 1984 was 1:30.000 and the minimum patch size that it could be interpreted was 1 ha. Ortophotos of 1998 had more resolution and minimum patch size decreased to 1000 m^2. Nevertheless, the landscape comparison between dates has been made at 1ha detail (Ortega et al., 2008). The patches are relatively homogeneous portions of land that represent different land covers that are adjacent and make up the landscape. The typology of land cover used derives from the CORINE land cover classification (EEA,1995) level 2 and is compatible with EUNIS habitat classification. When different land cover types were found in small patches (less than 1 ha) were considered "mosaic" land cover type. Additional codes related with land cover type composition were used to descript them.

SISPARES data are structured in a GIS database called SIGPARES which provide information per date concerning to land cover types, temporal change processes, aerial and field photographs and roads delineation. The shape files register information per patch relating to land cover type, canopy density, primary, secondary and thirty species, and other marginal landscape elements with their primary, secondary and thirty species, in the case of forest, "Dehesa" and "Matorral" land cover types. For forest plantation land cover type, manage and age are inventory. For crop land cover type, manage (irrigated or non irrigated) and typology (orchard, almond or vineyard, etc) are register. For water body and riparian woodland land cover types, the typology is inventory. Main land cover types division and localisation per plot are available in the web *http://www.chavales.net/index.html*.

8.3.2 Data management

The benchmark survey of SISPARES was the aerial photographs of 1984, in which areal, linear and point elements of landscape were delineated per plot. Comparative surveys were made with 1956 and 1998 aerial photographs in order to detect changes in some landscape elements. New shapes files were generated per date and the land cover type, canopy density and manage changes (1956 - 1984 and 1984 – 1998) were classified by change process and mapped. Landscape composition and configuration per plot and date were analyzed by means of structural indices and temporal changes mapped. Change process and landscape structural analysis are available in the web *http://www.chavales.net/index.html* .

8.3.3 Data transformation

There is a broad agreement between SISPARES land cover units and GHCs showed in the plot example of the deliverable D.4.1. of WP4 . The SISPARES additional codes including in the recording procedure (species composition, canopy density, manage, etc) can be used to further divide the principal division of SISPARES in GHCs. However, the data transformation of SISPARES monitoring in EBONE data (GHC) have a critical problem with the minimum mappable unit size as was reported in the deliverable D.4.1. of WP4 . SISPARES used 1 ha in 1956 and 1984 surveys and 1000 m2 in 1998 survey, whereas EBONE will use 400 m2. Nevertheless, based on the previous experience of S*ISPARES*, a methodological scheme has been proposed for adapting the approach to Spain (see Bunce et al 2006).

8.3.4 Data policy issues

Currently, there are no limits to use SIGPARES database inside EBONE project previous to a formally request.

8.4 British Country Side Survey / North Ireland Country Side Survey

The British Country Side Survey plots (n=265 (1978) and about 500 (from 1984) are selected using a stratified random selection procedure. They encounter for monitoring of species and habitat based on kilometre sample squares (1978, 1984, 1998, 2007); from 1984 on available in digital format (spatial data and database). The size of the kilometre sample squares is 1km² (1000x1000m). Data provided at the workshop were habitat mapping within the kilometre sample squares, n=10. Spatial data were missing.

In addition to the habitat recording also vegetation releveés are made. One vegetation plot in each habitat category (X, 10x10m), one on the edge of each category of linear features (G; 1x10m), one at each category of river ecosystems (S; 1x10m) and in addition (if needed) targeted plots in small scale Natura2000 habitat types (2x2m). No vegetation data were provided for the exercise.

The same accounts for the North Ireland Country Side Survey, n=2. Only different categories and a slightly different method is used. The size of the kilometre sample squares in 0.25km² (500x500m)

9 Outlook

A variety of reference models for integration of biodiversity data exist (e.g. LifeWatch, Alter-Net, Orchestra, etc.). For EBONE a simple and INSPIRE compliant architecture was chosen. The aim was to provide access to distributed data sources and to ensure a certain level of data harmonisation.

The described components will be tested in 2010 to provide a data management system for EBONE.

10 References

Adamescu, M., Cazacu, C., Peterseil, J., Datcu, Sabina., Schleidt, Kathi (2007). Report on LTER InfoBase. [Download 2009-02-02 from http://www5.umweltbundesamt.at/ALTERNet/index.php?title=Image:Report_LTER_InfoBase_version3_UNIBUC.zip]

Briesen, M., Hofmann, C., Otterstätter, A., Nikolai, R., Mirtl, M., Schentz, H., Vogel, W., Weber, K (2002). Dataintegration in the Field of Environmental Monitoring. In: Pillmann, W. & Tochtermann, K. (Eds) Environmental Communication in the Information Society; Proceedings of the 16th International Conference Informatics for Environmental Protection

Bunce, R.G.H., Roche, P., Bogers, M.M.B., Walczak, M., de Blust, G., Geijzendorffer, I., van den Borre, J. (2010). Draft Handbook for Surveillance and Monitoring of Habitats and Vegetation. Working Document for EBONE and BioBio. [Version 2010-01-28, download from https://www.surfgroepen.nl/sites/EBONE/Shared%20Documents/WP4/EBONE%20Handbook%20v20100128.pdf]. 98pp.

Darwin Core (2004). Darwin Core Schema (Version 1.3), A draft Standard of the Taxonomic Database Working Group (TDWG). [http://wiki.tdwg.org/DarwinCore].

Davenport, T. (1997). Knowledge management case study: Knowledge management at Microsoft.

Elkie, P.C., Rempel, R.S., Carr, A.P. (1999). Patch Analyst User's Manual. A Tool for Quantifying Landscape Structure. Ont. Min. Natur. Resour. Northwest Sci. & Technol. Thunder Bay, Ont. TM–002. 16 pp + Append.

Endbericht SINUS. http://131.130.59.133/projekte/sinus/sinus.htm (last accessed 10 Feb 2010)

GeoNetwork Community (2007). GeoNetwork OpenSource Complete Manual. [Version v2.4.0 2008-04-03, download from http://geonetwork-opensource.org/documentation/manual/geonetwork-manual/Manual.pdf] 152pp.

GeoServer (2010). GeoServer User Manual. Release 2.0.1. [Version 2.0.1 2010-01-11, download from http://downloads.sourceforge.net/project/GeoServer/GeoServer/2.0.1/GeoServer-2.0.1-pdfdoc.zip?use_mirror=heanet] 433pp.

Golfarelli, M., Rizzi, S. (2009). Data warehouse design: modern principles and methodologies. SEPS, Italy.

Gruber, T.R. (1993). Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli, (Eds.), International Workshop on Formal Ontology, Padova, Italy. Revised August 1993. International Journal of Human-Computer Studies, Volume 43 , Issue 5-6, pp. 907-928.

Guarino, N, Giaretta, P. (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: Mars, N.J.I. (Ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing. IOS Press. Pp. 25-32.

Hardisty, A. (2010). LifeWatch Technical Construction Plan European Strategies for Local Implementation. Deliverable 5.3.1. of EU Seventh Framework Programme (FP7) Infrastructures (INFRA-2007-2.2-01) " Life Watch e-Science and Technology infrastructures for biodiversity data and observatories" [download from http://www.lifewatch.eu/index.php?option=com_content&view=article&id=93&Itemid=18] . 29 pp.

Hernandez-Ernst, V., Poigné, A., Giddy, J., Hardisty, A., (2010). LifeWatch Technical Construction Plan – Reference Model. Deliverable 5.1.3 of EU Seventh Framework Programme (FP7) Infrastructures (INFRA-2007-2.2-01) " Life Watch e-Science and Technology infrastructures for biodiversity data and observatories" [download from http://www.lifewatch.eu/images/stories/PDFs/rm-v0.4.pdf]. 230 pp.

Inmon, W.H. (1995). Tech Topic: What is a Data Warehouse? Prism Solutions. Volume 1. 1995.

INSPIRE (2007). INSPIRE Metadata Implementation Rules. Technical Guidelines based on EN ISO 19115 and EN ISO 19119. [Version 1.1 2009-02-18, download from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf]. 74pp.

Jones, M.B., Berkley, C., Bojilova, J., Schildauer, M. (2001). Managing Scientific Metadata. IEEE Internet Computing 2001:59-68.

Lechtenbörger, J. (2001). Data warehouse Schema Design. DISBIS 79, Akademische Verlagsgesellschaft Aka GmhH.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F. (2009). An ontology for describing and synthesizing ecological observation data. Ecological informatics (in press).

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F. (2009). An ontology for describing and synthesizing ecological observation data. Ecological informatics (in press).

McCartney, PH, Jones, MB. (2002). Using XML-encoded metadata as a basis for advanced information systems for ecological research. In: Callaos N, Porter J, Rishe N, editors. Information systems development II. Proceedings from the 6th World Multi Conference on Systematics, Cybernetics and Informatics, Orlando, Fl: Internet Institute Information and System. p 379-84.

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G. (1997). Nongeospatial Metadata fort he Ecological Sciences. Ecological Applications 7(1):330-342.

Mirtl, M., Schentz, H. (1997). MORIS – An information system designed for longterm monitoring and comprehensive ecosystem research programmes; in: International Meeting for Specialists in Forest Ecosystems, Extended Abstracts. Forest Department of Bozen. Pp. 10-1997

Nativi, S., Leibovici, D., Vaccari, L., Diaz, L., Santoro, M. (2009). D.2.2.1. Report on requirements from GEOSS, INSPIRE, GMES, SEIS and related initiatives and projects. [Version 2009-12-23, download from //EuroGEOSS_D_2_2_1.doc]. 60pp.

Network Service Drafting Team (2008). INSPIRE Network Services Architecture. [Version 3.0 2008-07-19, download from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3_5_INSPIRE_NS_Architecture_v3-0.pdf]. 29pp.

OGC (2004). The Spatial Web. An Open GIS Consortium (OGC) White Paper. [Version 1 2004-06-14, download from http://portal.opengeospatial.org/files/?artifact_id=3859&version=2&format=pdf] 8pp.

OGC (2005). Web Service Implementation Specification. [Version 1.1.0 2005-05-03, download from http://portal.opengeospatial.org/files/?artifact_id=8339]. 131pp.

OGC (2007). Observations and Measurements – Part 1 – Observation schema. [Version 1.0 2007-12-08, download from http://portal.opengeospatial.org/files/?artifact_id=22466]. 85pp.

Peterseil, J. (2005). Bewertung der ökologischen Nachhaltigkeit in österreichischen Kulturlandschaften. Dissertation, Universität Wien.

Peterseil, J., Wrbka, T., Plutzar, C., Schmitzberger, I., Kiss, A., Szerencsits, E., Reiter, K., Schneider, W., Suppan, F., & Beissmann, H. (2004). Evaluating the ecological sustainability of Austrian agricultural landscapes: the SINUS approach. Land Use Policy, 21, 307-320.

Rajabifard, A., Kalantari, M., Binns, A. (2009). SDI and Metadata Entry and Updating Tools. In: B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Eds.) SDI Convergence. Research, Emerging Trends, and Critical Assessment. Netherland Geodetic Commission 48.

Renetzeder, C., Schindler, S., Peterseil, J., Prinz, M.A., Mücher, S., & Wrbka, T. (2010). Can we measure ecological sustainability? Landscape pattern as an indicator for naturalness and land use intensity at regional, national and European level. Ecological Indicators 10, 39-48.

Schentz, H., Mirtl, M. (2003). MORIS an universal information system for environmental monitoring. In: Schimak, G.P. & Swayne D.A. & Quinn N.W.T. & Denzer R. (Eds) Environmental Software Systems, Volume 5

Schentz, H., Schleidt, K., König, M., Mirtl, M. (2005). MORIS – MEDEA or using ecological tools for Monitoring or Meteorological Extreme Events. In: Hrebicek, J. & Racek, J. (Eds.) Informatics for Environmental Protection. Networking Environmental Information. Proceedings of the ENVIROINFO Brno 2005. Brno, Czech Republic. Pp. 365-370.

Schentz, H., Schleidt, K., Lane, M. (2006). Functional concept for a biodiversity and ecological network. (ALTER-Net) Project n. GOCE-CT-2003-505298. WP_I6_2006_04. 75 pp. [Download 2009-02-02 from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/ANet_WPI6_2006_04_Network_Concept.pdf]

Schentz, H., Schleidt, K., Lane, M. (2006). Functional concept for a biodiversity and ecological network. (ALTER-Net) Project n. GOCE-CT-2003-505298. WP_I6_2006_04. 75 pp. [Download 2009-02-02 from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/ANet_WPI6_2006_04_Network_Concept.pdf]

Schleidt, K., Lane, M., Schentz, H. (2006). Requirements for an IT Framework for ALTER-Net - Results from the Requirements Analysis conducted by Work Package I6. (ALTER-Net) Project n. GOCE-CT-2003-505298. WP_I6_2005_01. 21 pp. [Download 2009-02-02 from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/ANet_WPI6_2005_01_IT_Requirements.pdf].

Schleidt, K., Lane, M., Schentz, H. (2006). Requirements for an IT Framework for ALTER-Net - Results from the Requirements Analysis conducted by Work Package I6. (ALTER-Net) Project n. GOCE-CT-2003-505298. WP_I6_2005_01. 21 pp. [Download 2009-02-02 from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/ANet_WPI6_2005_01_IT_Requirements.pdf].

Schleidt, K., Schentz, H. (2006). Metadata Required for an ALTER-Net Data and Application Network. (ALTER-Net). Project n. GOCE-CT-2003-505298. WPI6_2006_03_Metadata_Report. [Download 2009-02-02 from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/ANet_WPI6_2006_03_Metadata_Report.pdf]

Schleidt, K., Schentz, H. (2006). Metadata Required for an ALTER-Net Data and Application Network. (ALTER-Net). Project n. GOCE-CT-2003-505298. WPI6_2006_03_Metadata_Report. [Download 2009-02-02 from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/ANet_WPI6_2006_03_Metadata_Report.pdf]

Sowa, J.F. (2000). Knowledge representation: logical, philosophical, and computational foundations. Brooks/Cole, Pacific Grove.

Van der Werf, B. Adamescu, M., Ayromlou, M., Bertrand, N., Borovec, J., Boussard, H., Cazacu, C., van Daele, T., Datcu, S., Frenzel, M., Hammen, V., Karasti, H., Kertesz, M., Kuitunen, P., Lane, M., Lieskovsky, J., Magagna, B., Peterseil, J., Rennie, S., Schentz, H., Schleidt, K., Tuominen, L. (2009). SERONTO. A Socio-Ecological Research and Observation Ontology: the core ontology. [Version 2009-04-09, Project n. GOCE-CT-2003-505298, download from http://www.alter-net.info/SITE/UPLOAD/DOCUMENT/outputs/WPI6_2009_10_SERONTOCore.pdf]. 26pp.

van der Werf, D.C., Adamescu, M., Ayromlou, M., Bertrand, N., Borovec, J., Boussard, H., Cazacu, C., van Daele, T., Datcu, S., Frenzel, M., Hammen, V., Karasti, H., Kertesz, M., Kuitunen, P., Lane, M., Lieskovsky, J., Magagna, B., Peterseil, J., Rennie, S., Schentz, H., Schleidt, K., Tuominen, L. (2008). SERONTO: a Socio-Ecological Research and Observation oNTOlogy. In: Weitzman, A.L., and Belbin, L. (Eds.). Proceedings of TDWG (2008), Fremantle, Australia.

Wrbka, T., Fink, M., Beissmann, H., Schneider, W., Reiter, K., Fussenegger, K., Suppan, F., Schmitzberger, I., P.uhringer, M., Kiss, A., Thurner, B., (2002). Kulturlandschaftsgliederung Österreich. Endbericht des gleichnamigen Forschungsprojektes. Forschungsprogramm Kulturlandschaft 13 (CD-ROM), Wien, Bundesministerium für Bildung, Wissenschaft und Kultur.

Zhuge, Y. Garcia-Molina, H., Wiener, J.L. (1996). The Strobe Algorithm for Multisource warehouse consistency. In *Proccedings Conference on Parallel and Distributed Information Systems*, Miami.

ALTERRA WAGENINGEN UR

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

The University of Edinburgh

CSIR
our future through science

SLU
Sveriges lantbruksuniversitet

ile
Institute of Landscape Ecology
Slovak Academy of Sciences

HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

NINA

Cemagref

Universidad Politécnica Madrid

רשות הטבע והגנים

inbo

Facultatea de Biologie
Universitatea din Bucureşti

umweltbundesamt

JRC
EUROPEAN COMMISSION

universität wien

Eesti Maaülikool
EMÜ Estonian University of Life Sciences

ARISTOTLE UNIVERSITY OF THESSALONIKI