BIO_SOS

Project Title:	**BIO_SOS Biodiversity Multisource Monitoring System: from Space TO Species**
Contract No:	FP7-SPA-2010-1-263435
Instrument:	
Thematic Priority:	
Start of project:	1 December 2010
Duration:	36 months

Accompanying report to Deliverable No: D4.5

(Collaborative platform for data sharing)

The BIO_SOS metadata geoportal and the external quality of pre-existing datasets

Due date of deliverable:	November 30, 2011
Actual submission date:	December 14, 2011
Version:	1st version of D4.5
Main Authors:	João Pradinho Honrado, Joaquim Alonso, Pedro Castro, Luís Martins, Isabel Pôças, João Gonçalves, Carlos Guerra, Bruno Marcos

Project ref. number	**263435**
Project title	**BIO_SOS: Biodiversity Multisource Monitoring System: from Space to Species**

Deliverable title	**The BIO_SOS metadata geoportal and the external quality of pre-existing datasets**
Deliverable number	Del no D4.5
Deliverable version	1st version of D4.5
Previous version(s)	n/a
Contractual date of delivery	November 30, 2011
Actual date of delivery	December 15, 2011
Deliverable filename	BIO_SOS_D4.5_v01
Nature of deliverable	R = Accompanying Report to D4.5
Dissemination level	PU = Public
Number of pages	123
Workpackage	WP4
Partner responsible	Partner 9 (CIBIO/ICETA)
Author(s)	João Pradinho Honrado, Joaquim Alonso, Pedro Castro, Luís Martins, Isabel Pôças, João Gonçalves, Carlos Guerra, Bruno Marcos
Editor	João Pradinho Honrado
EC Project Officer	Florence Beroud

COOPERATION

Abstract	As a synthesis of work developed in Task 4.1 (months 3-12), this report provides (i) a detailed description of the requirements and technical specifications of the new BIO_SOS WebGIS platform and metadata geoportal, and (ii) an evaluation of the external quality (fitness-for-use and fitness-for-purpose) of core pre-existing datasets, based on a quantitative, indicator-based comparison of internal quality (based in metadata) and user requirements and expectations. Finally, the report describes an implementation and development plan for the geoportal and recommendations for data quality evaluation and management throughout the project. The BIO_SOS WebGIS platform and metadata geoportal described in this report can be accessed at http://81.90.50.40/ (provisional link).
Keywords	External quality evaluation, Metadata geoportal, Pre-existing datasets, WebGIS platform.

Signatures

Written by	**Responsibility- Company**	**Date**	**Signature**
João Pradinho Honrado	Responsible of WP4, CIBIO	14/12/2011	
Verified and approved by			
Jens Stutte	As Responsible of WP3 and Quality Manager (temporal substitute), PKI	14/12/2011	
Approved by			
Palma Blonda	Coordinator, CNR	15/12/2011	

Table of Contents

1. Executive summary

This report is related to **deliverable D4.5** and in this sense it provides a synthesis of the work carried out in Task 4.1, within work package (WP) n. 4 of BIO_SOS. It is joined by a newly developed metadata geoportal, supported by a collaborative platform (accessible at http://81.90.50.40/), aimed at facilitating data management, sharing and quality evaluation throughout the project. Together with D4.1, this deliverable also relates to milestone MS1, i.e. to the identification and harmonized description of pre-existing datasets for training and test sites. This report therefore completes the previously delivered D4.1, which provided a synthesis on a first inventory and internal quality analysis of pre-existing datasets within the consortium, as well as a proposal of a methodological framework for external quality evaluation of pre-existing and newly produced datasets. The workflow in the second stage of Task 4.1 (Months 9-12) included the final specification and implementation of a metadata geoportal and the implementation of the methodological framework for quality evaluation of spatial datasets.

The overarching rationale behind the development of Task 4.1 was based on the fact that **external quality assessment** is, by definition, a user-oriented process and should be based on a quantitative, indicator-based comparison of internal quality (based on metadata) and user requirements and expectations. These are determined by intended service outputs and their quality, which in the case of BIO_SOS should mean a close involvement of final Users (i.e. agencies and other stakeholders) in all tasks across the project involving data quality evaluation. The quality evaluation routines implemented in the new BIO_SOS metadata geoportal are expected to facilitate and harmonize such tasks across the consortium and to support the quality expectations of project Partners and of a range of potential end-users. In this context, in this report the following contents can be found: (i) an introduction to data discovery, sharing and communication in the scope of BIO_SOS (Section 2); (ii) the definition of platform and users' requirements for a metadata geoportal within BIO_SOS (Section 3); (iii) a detailed description of the specifications of the new metadata geoportal (Section 4); (iv) the evaluation of external quality of core pre-existing datasets according to several application contexts relevant for project objectives and tasks (Section 5); (v) a proposal of a development and implementation plan for the geoportal throughout the project (Section 6); and finally (vi) a synthesis with guidelines and recommendations for data quality management across the project (Section 7).

The **BIO_SOS metadata geoportal** represents a key step towards data transparency and sharing within the project consortium. This is a critical issue not only to the harmonization of data organization procedures towards a framework of data and system interoperability, but also to the establishment of workflows within the project in order to meet its development objectives and main goals. In this context, the geoportal requirements include:

(i) Metadata editing tools, capable of supporting an INSPIRE compliant metadata profile;

(ii) Validation tools and procedures that can enforce quality standards at the level of metadata editing;

(iii) Discovery services and tools that allow searching, both thematically and spatially, the available data within the consortium; and (iv) quality evaluation tools and procedures that allow to infer about the overall quality of the datasets and databases for specific uses and modelling processes.

The metadata geoportal has been developed as a **collaborative platform** with functionalities to catalogue all relevant in situ and ancillary data across the consortium, based on a common metadata profile promoting the harmonization of dataset description, to promote the sharing, searching and retrieval of metadata among project Partners, as well as to facilitate the visualization of reference and thematic datasets, in both raster and vector formats, besides creating a platform to support the development of tools for data quality assessment based on metadata. The developed metadata geoportal provides a simple and easy-to-use graphical user interface, only accessible to authorized users, consisting of three distinct parts which are accessible through a web page address: the geoviewer inter-

face, the metadata management interface, and the user management interface. The geoviewer interface provides an interactive map that allows the visualization and visibility control of several reference and thematic cartographic data, whereas the metadata management component focuses on the searching, access and editing of metadata entries. Finally, the user management component allows the definition of groups, users and their privileges regarding access to metadata entries. Additionally, the system integrates catalogue services to allow metadata searching and retrieving to facilitate interoperability and integration with internal and external information systems.

The methodological framework implemented for **external quality evaluation** (EQE) is based on an end-user, objective-oriented and data-centric perspective, meaning that the external quality of datasets is assessed for each application context by measuring the overall matching between the intrinsic characteristics of the data (detailed by the BIO_SOS standard metadata profile) and the characteristics of the data as required by users. This EQE framework was applied to the core pre-existing datasets identified by each site Partner, representing the most relevant datasets for a specific geographical area and for a set of specified application contexts. Several case studies reporting results of EQE for core pre-existing datasets identified per application context and site/country are described. In general, those results revealed a medium to high average value of fitness for use for the studied set of core datasets. The results also emphasize the importance of an accurate filling of metadata information, which should also be a concern for new datasets to be acquired or produced throughout the project. Furthermore, besides the core pre-existing datasets identified by Partners, other datasets (both pre-existing and newly acquired or produced) may be necessary to support different application contexts throughout the project, and their EQE must be timely performed. Therefore, the continuous and iterative character of quality evaluation is once again stressed, towards the improvement and validation of datasets for specific purposes.

The development of the WebGIS platform within the time frame of Task 4.1 was influenced by two specific processes/phases that constrained its progress, namely:

(i) The collection of metadata from all Partners in charge of training and test sites;

(ii) The definition of the methodologies as well as the design and implementation of the technologies that were used.

The **future development** process of the BIO_SOS WebGIS platform and metadata geoportal will include:

(i) The integration of new components/modules in the platform (e.g. the external quality evaluation module);

(ii) The loading of metadata for newly acquired or produced datasets (to support future (meta)data sharing within the consortium);

(iii) The assessment of platform functioning by team members of all Partners in the consortium.

The implementation plan beyond the delivery of D4.5 (end of month 12), running at least until month 20 of the project, intends to optimize the BIO_SOS WebGIS platform and metadata geoportal. Thus, the exploration and optimization of the results and the continuity and sustainability of the platform should involve governance strategies and practices at the different levels and components (data, technologies, users, standards and policies).

Now at its completion, **Task 4.1 (deliverables D4.1 and D4.5) has contributed to**:

(i) Evaluate, select, organize and share (through metadata) relevant and potentially useful pre-existing datasets;

(ii) Identify important data gaps and establish priorities for new dataset acquisition;

(iii) Establish guidelines and tools for data quality evaluation and management across the project;

(iv) Develop a collaborative platform for (meta)data sharing within the consortium.

Particularly, the data quality evaluation procedures established in Task 4.1, and implemented in the geoportal, are of high importance for future work in BIO_SOS, since, among other reasons: (i) they provide an evaluation of pre-existing datasets in their quality and relevance for the several WPs and Tasks of BIO_SOS; (ii) they signal important data gaps with potential consequences for the workflow in the project; (iii) they contribute to identify opportunities (and limitations) for studies of specific processes of change across sites; and (iv) they may contribute to the establishment of a methodological reference ("best practice") for similar projects. Beyond the specific context and objectives of BIO_SOS, this quality evaluation of pre-existing datasets across several European and non-European countries will provide a formal assessment of the actual usefulness of a wide range of habitat, biodiversity and ancillary datasets to support or frame the monitoring of habitats, biodiversity and landscapes under international goals, targets and indicators.

2. INTRODUCTION

2.1 Scope of the report

This report is related to deliverable D4.5 ("The BIO_SOS metadata geoportal and the external quality of pre-existing datasets") and in this sense it provides a synthesis of the work carried out in Task 4.1 ("Collection and analysis of pre-existing data"), within work package n. 4 (WP4, "On-site data collection") of BIO_SOS. It is joined by a newly developed metadata geoportal, supported by a collaborative platform, aimed at facilitating data management, sharing and quality evaluation throughout the project. This report completes the previously delivered D4.1 ("Report on pre-existing in situ and ancillary datasets for sites"), which provided a synthesis on a first inventory and internal quality analysis of pre-existing datasets within the consortium, as well as a proposal of a methodological framework for external quality evaluation of pre-existing and newly produced datasets regarding their potential relevance and usefulness for the objectives and tasks in the project.

In this context, in the several sections of the report the following contents can be found: (i) an introduction to data discovery, sharing and communication in the scope of BIO_SOS (Section 2); (ii) the definition of platform and users' requirements for a metadata geoportal within BIO_SOS (Section 3); (iii) a detailed description of the specifications of the new metadata geoportal (Section 4); (iv) the evaluation of external quality of core pre-existing datasets according to several application contexts relevant for project objectives and tasks (Section 5); (v) a proposal of a development and implementation plan for the geoportal throughout the project (Section 6); and finally (vi) a synthesis with guidelines and recommendations for data quality management across the project (Section 7).

As described in D4.1, the overarching rationale behind this organization of work in Task 4.1 is based on the fact that external quality assessment is, by definition, a user-oriented process and should be based on a quantitative, indicator-based comparison of internal quality and user requirements and expectations (see Section 5). These are determined by intended service outputs and their quality, which in the case of BIO_SOS should mean a close involvement of final Users (i.e. agencies and other stakeholders) in all tasks across the project involving data quality evaluation, as a follow-up of the signed Service Level Agreements (see deliverable D2.3). The quality evaluation routines implemented in the BIO_SOS metadata geoportal are expected to facilitate and harmonize such tasks across the consortium and to support the quality expectations of project partners and of a range of potential end-users.

2.2 Data discovery, sharing and communication in the scope of BIO_SOS

The BIO_SOS WebGIS collaborative platform intends to be a major contribution to strength the communication channels between Partners and to establish a collaboration framework based on a (meta)data and knowledge sharing network. In this sense, the BIO_SOS WebGIS collaborative platform aim is to integrate different open source applications in order to capture, search and share spatial data and data services, including a metadata geoportal that implements a standard web catalogue service. Within this scope, the BIO_SOS metadata geoportal was implemented in order to organize and describe all datasets in the BIO_SOS consortium. Moreover, a quality evaluation standard was defined and (at this stage) partially implemented to establish quality evaluation routines for pre-existing and, more important, for future datasets and process modeling to be developed within the project.

The metadata geoportal will allow: (i) to upload and manage metadata from the different training and test sites of the project, including pre-existing datasets and newly collected or derived datasets; (ii) to consult, discover, access and edit the metadata; (iii) through the implementation of CSW standards, to communicate and share metadata with the EODHaM system, as well as with other internal (and external) instruments of file transfer, e.g. FTP

platforms implemented in BIO_SOS and in other projects (see deliverable **D4.2**); and (iv) to perform quality evaluations of the available databases. Given its flexibility and collaborative environment, the implemented tools and management strategies will promote platform sustainability according to other BIO_SOS final products. In the future, this WebGIS platform should act as a prime instrument to support quality evaluation of pre-existing and new datasets and as a core mechanism for internal and external communication.

Therefore, in the broader scope of BIO_SOS, the WebGIS platform, and particularly the metadata geoportal, will greatly contribute to: (i) define cross country/partner modeling procedures; (ii) implement quality evaluation routines; (iii) define and enforce communication channels between partners, but also externally with other projects; and (iv) provide a data discovery and sharing framework for institutions involved in the consortium.

2.3 D4.5 within the broader scope of BIO_SOS

In the broader context of WP4 objectives within BIO_SOS ("collecting, harmonizing and sharing pre-existing datasets on sites relevant for habitat mapping, and supplementing existent datasets with new field data from on-site campaigns based on standard protocols"), **Task 4.1** intends to: (i) identify datasets, projects and institutional data providers; (ii) describe and collect all relevant in situ and ancillary data from the several countries; (iii) organize and harmonize all datasets on common standards; and (iv) provide a collaborative platform to catalogue, query and share databases among project partners using an internal network, particularly to feed other WPs as well as other tasks in WP4.

As described in detail in later sections of this report, the workflow in the second stage of Task 4.1 (Months 9-12) included the final specification and implementation of a metadata geoportal and the implementation of the methodological framework for quality evaluation of spatial datasets (described in **D4.1**), harmonized with the general timeline defined for this task in the Description of Work. Therefore, the overall workflow of Task 4.1 (Months 3-12) started with the collection of simplified metadata on all pre-existing datasets concerning training and test sites, and is completed with the implementation of a collaborative platform to support metadata sharing among partners and data quality evaluation within the project (Figure 2.1).

		WP4																								
	Months	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	Pre-existing datasets						D41			MS1	D45															
2	Criteria for EO selection			D44																						
3	Ongoing projects								D42																	
4	On site data collection						D43																MS2			

Milestone MS1 - Pre-existing data have been collected and harmonized for all test sites
Milestone MS2 - Data from new on-site field campaigns have been collected from all sites
Deliverable D41 - Report on pre-existing in situ and ancillary datasets for sites
Deliverable D42 - Report of the connection to other projects
Deliverable D43 - Report on protocols for new on-site in-field campaigns
Deliverable D44 - Report on criteria for selection of suitable EO datasets and identification of EO datasets with adequate range of spectral, spatial and temporal resolutions for each site
Deliverable D45 - Database and collaborative platform for sharing pre-existing data

Figure 2.1 - D4.5 within the general timeline of Task 4.1 and relations with milestones and other deliverables within the Task and WP4.

In this sense, D4.5 relates with several other deliverables in WP4 (Figure 2.1), since: (i) as a follow-up of **D4.1**, it provides the results of the external quality evaluation of core pre-existing datasets as well as the specifications of the newly developed metadata geoportal; (ii) it complements D4.4 (Task 4.2) in the assessment of data needs and requirements, as well as in the evaluation of relevance of pre-existing datasets; and (iii) together with D4.4 and D4.5, it helps identifying data gaps and establishing priorities for new data acquisition, either through financial investment (e.g. new very-high resolution imagery), field campaigns (D4.3, Task 4.4), or connection to other projects and initiatives (D4.2, Task 4.3). Together with **D4.1**, this deliverable also relates to milestone MS1 ("Pre-existing data have been collected and harmonized for all test site", Month 11; Figure 2.1), more specifically to the identification and harmonized description of pre-existing datasets for training and test sites.

D4.5 also relates with on-going and future work in the broader scope of BIO_SOS. In fact, provided they fit the pre-established quality requirements, pre-existing datasets have been, and will be valuable, in several stages of BIO_SOS, namely: (i) in the identification and selection of key processes and drivers of ecological change in each site (WP2); (ii) in the selection of focal areas within sites for EO imagery selection and acquisition (WP4: Task 4.2); (iii) in the identification of crucial data gaps and selection of key on-going projects which may provide important datasets (WP4: Task 4.3); (iv) in the support to sampling designs for new on-site campaigns to collect data on habitats, biodiversity or pressures (Task 4.4); (v) in the support to EO image classification (WP5) and final validation (WP7); and (vi) in the modelling of relations between EO data, habitat classifications, landscape patterns, and focal indicators adopted in BIO_SOS (WP6). In this context, **D4.5** (as a follow-up of **D4.1**) relates with several milestones and deliverables from other WPs in BIO_SOS, since: (i) it builds upon previous work on indicators (D2.1), sites and pressures (D2.2), user requirements (D2.3) and service design (D3.1) in WP2 and WP3; and (ii) together with previous and on-going deliverables from WP5 and WP6, it provides results which are important for future work in WP5, WP6 and WP7.

Now at its completion, Task 4.1 (including deliverables **D4.1** and **D4.5**) has contributed to: (i) evaluate, select, organize and share (through metadata) relevant and potentially useful pre-existing datasets; (ii) identify important data gaps and establish priorities for new dataset acquisition; (iii) establish guidelines and tools for data quality evaluation and management across the project; and (iv) develop a collaborative platform for (meta)data sharing within the consortium. Particularly, the data quality evaluation procedures established in Task 4.1, and implemented in the geoportal, are of high importance for future work in BIO_SOS, since, among other reasons: (i) they provide an evaluation of pre-existing datasets in their quality and relevance for the several WPs and Tasks of BIO_SOS; (ii) they signal important data gaps with potential consequences for the workflow in the project; (iii) they contribute to identify opportunities (and limitations) for studies of specific processes of change across sites; and (iv) they may contribute to the establishment of a methodological reference ("best practice") for similar projects. Beyond the specific context and objectives of BIO_SOS, this quality evaluation of pre-existing datasets across several European and non-European countries has provided a formal assessment of the actual usefulness of a wide range of habitat, biodiversity and ancillary datasets to support or frame the monitoring of habitats, biodiversity and landscapes under international goals, targets and indicators.

3. REQUIREMENTS FOR A BIO_SOS METADATA GEOPORTAL

3.1 Platform requirements within the project

The BIO_SOS WebGIS platform intends to support internal communication within the consortium, and simultaneously, with the definition of the project products and their dissemination strategy and instruments, to contribute to an external communication strategy with other projects in the thematic scope of BIO_SOS. The collaborative nature of the platform is ensured by the broad access and contributions from all Partners at the level of content production, functioning and management. In this sense, it is expected that the development and integration of different platform technological components ensures the completion of current requirements, management and sharing of metadata, but also the platform development and maintaining continuity according to a modular, scalar and evolutionary implementation of the BIO_SOS platform. The provisional dimension of the BIO_SOS project and the associated costs contributed to a choice for an implementation strategy based on open-source software.

The requirements and platform specifications are critical for the development, implementation and testing of the BIO_SOS WebGIS platform and metadata geoportal, particularly to improve users' adoption of the platform and its effective continuity. The development of the platform implied: (i) consulting and analyzing projects and WebGIS platforms with similar objectives and technological solutions; (ii) analyzing metadata management and production according to the specific legal and normative framework; (iii) thoroughly discussing, within the technical team in charge of Task 4.1 and deliverable **D4.5**, all issues regarding requirements, contents and functions; as well as (iv) considering and testing different technologies and development methodologies.

The WebGIS platform development process included: (i) the definition and characterization of the users and related access profiles; (ii) the definition and development of a metadata profile and of a metadata editor; (iii) the collection, capture and loading of metadata according to standardized tools and methodologies; (iv) the generation and management of metadata catalogues; (v) the selection of relevant datasets and assessment of their quality using internal and external quality evaluation tools based on metadata; and (vi) the conceptualization, development and implementation of a graphical interface, as well as of a geographic viewer and thematic and spatial queries associated to the functionalities of the metadata geoportal .

Furthermore, future advances of the BIO_SOS WebGIS platform will meet a set of specific requirements: (i) in functional interaction with other internal communication technologies (e.g. the FTP platform) and datasets derived from image processing and habitat mapping (EODHaM); and (ii) potential data sharing with other technological platforms, information systems and spatial data infrastructures. In the first case, the platform will allow to search, discovery and locate resources (relevant datasets) of high importance for effective internal sharing using the FTP platform. EODHaM needs will also be supported by the identification and location of available datasets and associated external quality indicators. The new datasets to be produced in the project (namely land cover/use and habitat maps) should be included as new entries in the WebGIS platform through the complete filling of the established metadata profile. Therefore, a close functional interaction of the several platforms is expected, in compliance with the policies and data-sharing practices adopted within the project. Finally, the adoption of standards, guidelines and technical references will enable the future implementation of geowebservices with other platforms (including those external to the project) in order to establish communication networks and promote knowledge sharing.

3.2 User types and requirements

User types and their requirements determine not only the hierarchy of potential uses but also several strategic issues regarding the development and implementation of the metadata geoportal. Some of these issues are related to the necessary security and data warehousing protocols, in order to guarantee the quality and cross-platform interoperability of the collected metadata, as well as the implementation of an user structure compatible with the consortium expectations, i.e. taking into consideration the existence of several Partners with autonomous data production procedures.

In this context, the development of the BIO_SOS metadata web platform is based on a user (and use) oriented interface, which allows users to collect, systematize and publish metadata on pre-existing and newly produced spatial information within the consortium. In the metadata geoportal, five types of user profiles could be identified and characterized (Figure 3.1): (i) registered user (conducting search and visualization); (ii) editor (conducting search, visualization and edition); (iii) reviewer (reviewing metadata inputs and authorizing their publishing); (iv) user administrator (performing new users creation and administration within the own group); and (v) administrator (performing the administration of users and of the platform itself).

Based on these user profiles, a set of standard requirements was determined for all users, namely the need for: (i) tools and/or interfaces for contextual data visualization, based on a geographic visualization interface with a set of selected thematic layers that provide a contextual view for all sites (see 4.2.1), (ii) discovery services and user customized search tools, that allow users to identify and confirm the existence of specific datasets in each Partner's database and across the consortium; and finally (iii) database quality evaluation, that, based on previously defined quality indicators (see Section 5.2.3), will allow users to determine basic quality traits for the available datasets, as well as to perform data (external) quality evaluation for their own application contexts. Beyond these three standard and overarching requirements, the distinct user profiles will have different demands according to their specific requirements:

i) **Administrator Profile**:

a. editing capabilities;
b. metadata quality check routines and tools;
c. user managements tools; and
d. platform management tools.

ii) **Reviewer Profile**:

a. metadata quality check routines and tools; and
b. metadata reviewing and acceptance tools.

iii) **Editor Profile**:

a. editing capabilities; and
b. metadata quality check routines and tools.

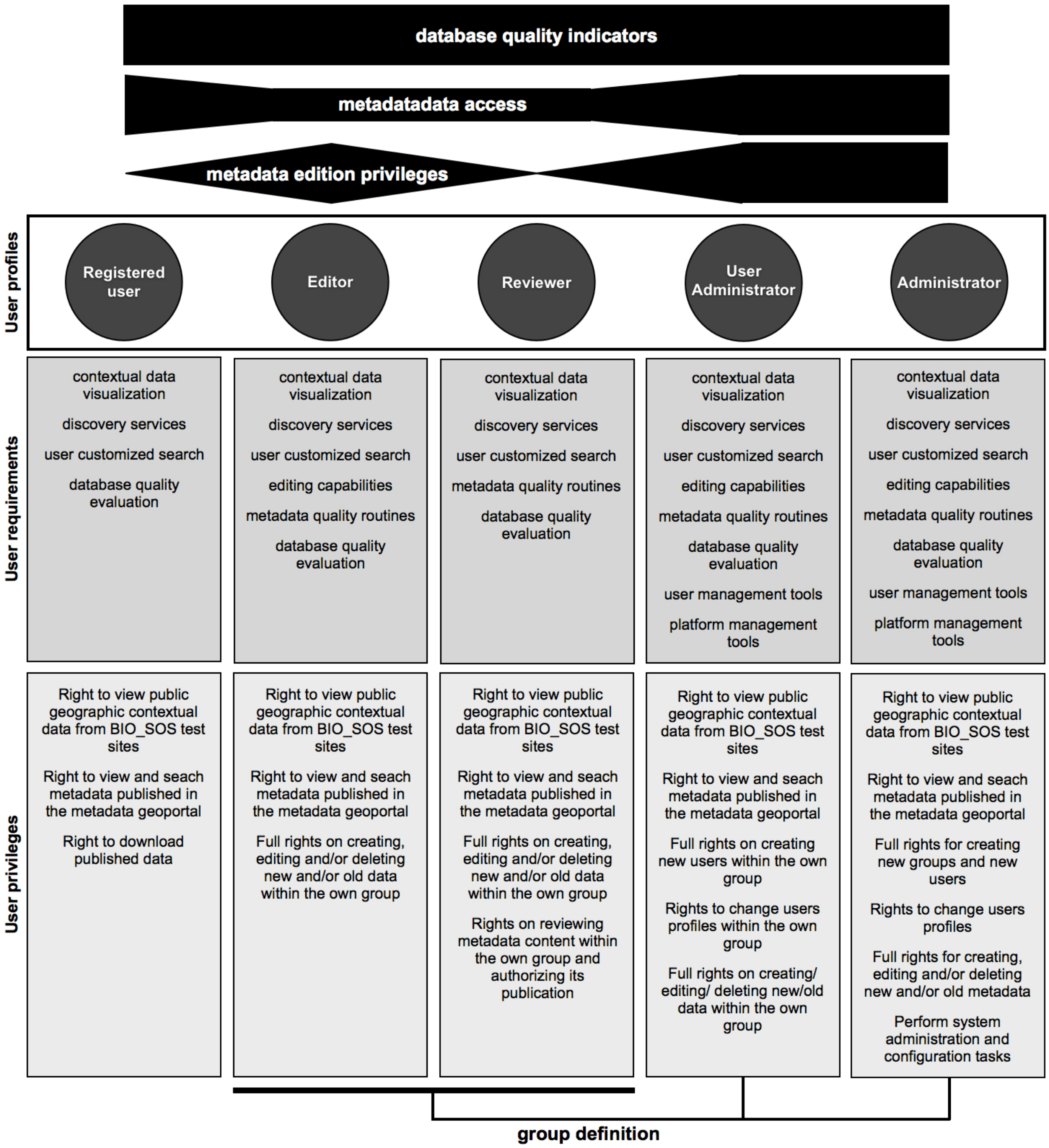


Figure 3.1 User profiles, their requirements and privileges in the context of the BIO_SOS metadata geoportal.

User privileges are defined according to the identified users' requirements but also, due to software constraints, according to a predefined set of user profile rights. In this context, user privileges for each user profile are listed below (adapted from GeoNetwork open-source, 2011a) and described in Figure 3.1:

i) The **Administrator Profile** has special privileges that give access to all available functions. These include: full rights for creating new groups and new users; rights to change users'/groups' profiles; full rights for creating, editing and/or deleting new and/or old metadata; view and search metadata published in the metadata geoportal; and perform system administration and configuration tasks.

ii) The **user Administrator Profile** has privileges to create new users within his own group. The privileges also include rights to change users profiles within the own group and full rights on creating/editing/deleting new/old data within the own group, and to view and search metadata published in the groups to which he/she belongs in the metadata geoportal.

iii) The content **Reviewer Profile** is meant to give final clearance for metadata publication on the Intranet (private view) and/or on the Internet (public view). This user profile has to be present in each group of users to be created, and its privileges include: rights on reviewing metadata content within the own group and authorizing its publication; full rights on creating/editing/deleting new/old data within his/her own group; and view and search metadata published in the groups to which he/she belongs in the metadata geoportal.

iv) The **Editor Profile** corresponds to the main metadata edition profile, and users with this type of profile have privileges that include: full rights on creating, editing and/or deleting new and/or old data within their own group(s); and view and search metadata published in the groups to which he/she belongs in the metadata geoportal.

v) The **Registered User Profile** has restricted access privileges to the metadata uploaded into the metadata geoportal. These restrictions reflect mainly on the absence of edition privileges and allow this type of user to only be able to view and search metadata published in the groups to which he/she belongs in the metadata geoportal.

It is important to highlight that these different user profiles have differentiated access to the metadata, as only the platform administrator profile has access to all the published content, while the reviewer, the editor and registered user profiles only have access to the published metadata within their own groups. These different privileges determine that a specific user may have to belong to two different groups in order to be able to have a wider access to metadata entries within the context of the metadata geoportal.

3.3 Platform functionalities and utilization requirements

The implementation of the BIO_SOS metadata geoportal has three main domains: (i) the first one is related to its intrinsic functions; (ii) the second one relates to the interaction with other implementation tasks/work packages in the project; and (iii) the third one is related to the implementation of interoperability towards efficiently sharing information (metadata) with other catalogue services outside the consortium. Within this scope of application, platform utilization requirements were defined having in consideration the specific traits of each domain, and also their main objectives:

i) related to intrinsic functions:

a. provide a catalogue service for metadata within the consortium;

b. provide a clear view of the available datasets within the consortium;

c. provide search and discovery tools;

d. establish a standard for data harmonization;

e. promote transparency in data management and utilization;

ii) within the consortium:

a. provide the means to catalogue new data within the consortium;

b. provide metadata-based evaluation tools for data quality and availability assessment;

c. improve communication within the consortium;

iii) outside the consortium:

a. implement communication protocols that allow interoperability with other catalogues;

b. contribute to project communication outside the consortium;

Considering these objectives and the number of Partners involved in the consortium, the WebGIS platform should provide: (i) discovery and access services based on communication protocols/standards that allow communication services with other platforms (namely CSW OGC standards [Web Catalogue Service]); (ii) metadata oriented search tools; (iii) user management tools; (iv) data quality evaluation protocols and tools; and (v) a communication interface with a public access to improve communication within and outside the project consortium. Discovery and access services are at the core of the entire platform as they allow an interoperable basis to all contexts and domains of application/implementation within and outside the consortium. CSW OGC standards [Web Catalogue Service] will be implemented in order to guarantee the existence of communication standards easily recognized by other catalogue services available worldwide, but also to anticipate the necessity of implementing interoperable metadata tools (from the point of view of the metadata geoportal) in the future EODHaM service.

Another critical issue associated to the metadata geoportal is the need to implement a search engine oriented to search for metadata, and ultimately datasets, available within the consortium. This will allow assessing the existence of the necessary datasets to instruct a specific analytical or modelling process, but also identifying possible collaboration niches inside the partnership. These may represent two major advances in improving data analysis capabilities and cooperation within the BIO_SOS consortium. In articulation with this last feature, quality evaluation tools will help to determine if the available datasets have the necessary quality to produce the expected results, but also to determine (or support the decision on) the processes and/or expected results that a Partner can foresee with the available datasets, focusing on pursuing the most favourable application domains.

4. SPECIFICATION OF THE BIO_SOS METADATA GEOPORTAL

4.1 Logical and technological WebGIS architecture

The BIO_SOS metadata geoportal (accessible at http://81.90.50.40/) has been developed as a multi-platform tool to facilitate the metadata management, the sharing and visualization of relevant in situ and ancillary data among project Partners using an internal network, and enabling real-time data sharing capabilities, besides creating a platform to support the development of (semi-)automated tools for data quality assessment based on metadata. According to the objectives of the project, the following features, functionalities and constraints were considered in the context of system development: (i) the geoportal must promote the access, use and sharing of metadata information among project Partners through the World Wide Web; (ii) the use of Free Open Source Software should be preferred for the implementation of the portal in order to provide a low-cost development and maintenance solution; (iii) the system must provide simple and easy-to-use graphical user interfaces for creation, entry, and update of metadata information, since most of the system's users may have no prior experience on metadata handling; (iv) the platform must include the tools to record and manage alphanumeric and spatial data in an unique and integrated system, as well as to manage users' profiles and access privileges; and (v) the system must integrate catalogue services to allow metadata searching and retrieving in order to facilitate the interoperability and integration with other internal and external information systems.

Considering these requirements and, in order to create a modular based system to facilitate the maintenance and future development of new functionalities, the BIO_SOS metadata geoportal has been structured following a multi-layer client-server architecture model composed by the integration of different selected/developed components. This approach permits the creation of a flexible platform allowing developers to add, modify or upgrade specific components as the platform evolve.

The BIO_SOS metadata geoportal employs an architecture with three layers: (i) a data storage layer; (ii) a data access layer, and; (iii) a presentation layer (figure 4.1). The data storage layer includes the files and databases to store metadata information, spatial data, user profiles, and configuration data of the geoportal services and of the presentation layer. The data access layer comprises the data access and processing functionalities of the system, and consists of components to provide: the management and administration of the platform and users information; the creation and updating of datasets metadata; the capacity to query and retrieve contents of the metadata database; and services for map viewing. The presentation layer is responsible for the user interaction and integrates components to create a graphical user interface accessible through a Web browser, which provides functionalities for user administration, metadata searching and visualization (text and geographic), and metadata edition.

In order to implement the multi-layer architecture of the system, several software components were selected to fulfill the functionalities defined for the BIO_SOS metadata geoportal. In the software selection decision process, several factors were considered such as fulfillment of functionalities, knowledge and experience acquired in previous projects, compliance with standards, license type, easiness of installation, configuration and usability, existence of support resources, size of users community and availability of online documentation and examples.

With respect to the data storage layer, PostgreSQL database with the PostGIS spatial extension was selected as the database management system (DBMS) for spatial vectorial data storage and as GeoNetwork database to store datasets metadata and users profiles and permissions. PostgresSQL (PGDG, 2011) is an open-source object-relational database system providing the features of advanced DBMS such as complex query execution, triggers, stored procedures, multiuser access, transactional integrity and multi-version concurrency control. PostGIS (PostGIS, 2011) is an extension compliant with the OpenGIS Simple Feature Specification for SQL (OGC, 2010) that adds support for geographic objects to the PostgregSQL so it can be used as a spatial database for geographic

information systems. Regarding spatial raster data, a file-based approach was used since spatial raster data support in PostGIS is an ongoing project.

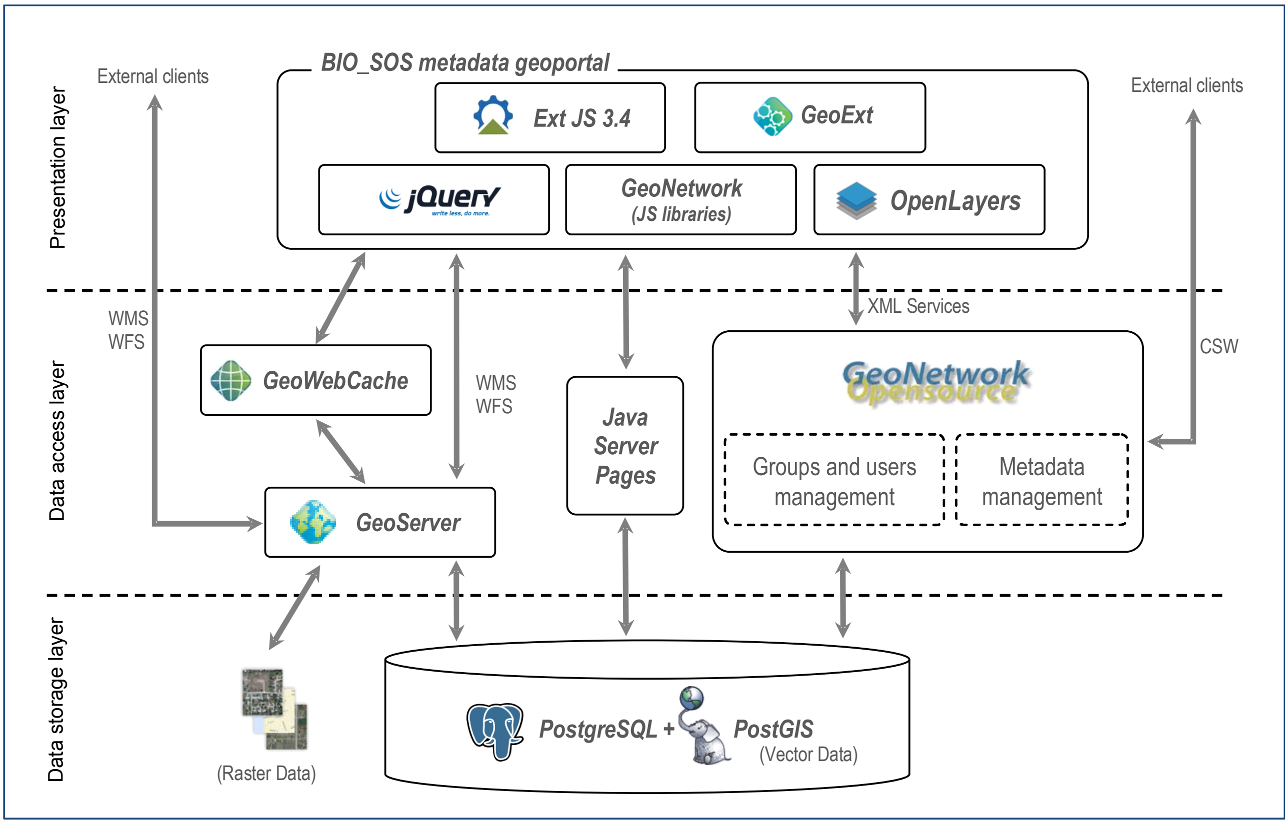


Figure 4.1 – General architecture of the BIO_SOS metadata geoportal.

Concerning the data access layer, two main software were considered to fulfill the necessary functionalities: Geoserver to publish spatial data, and GeoNetwork to carry out tasks of metadata management and users administration. GeoServer (GeoServer, 2011) is an open source software server that allows to view, share and edit geospatial data. It allows to publish spatial information from a wide range of data source for both vector and raster data in compliance with the Open Geospatial Consortium (OGC) standards, such as the Web Map Service (WMS), the Web Feature Service (WFS), and the Web Coverage Service (WCS). On another hand, GeoNetwork (GeoNetwork Opensource, 2011a) is an open source cataloging application to enable the management of geographically referenced resources, allowing to share spatially referenced thematic information. It includes a broad set of features, such as online metadata editing, geospatial data search across multiple catalogues, metadata harvesting, and synchronization of metadata between distributed catalogues. GeoNetwork implements the OGC Catalogue Services for the Web (CSW) specification (OGC, 2007) and is compliant with the international standards ISO 19115 "Geographic Information - Metadata" and ISO 19139 "Geographic information - Metadata - XML schema implementation". Additionally, the GeoNetwork software includes the option to manage metadata access through the management of groups' and users' accounts and permissions. In the context of the BIO_SOS metadata platform, GeoNetwork is used as a server to manage metadata, to govern users' authentication and access control, and as a CSW server.

Finally, the top layer of the architecture, the presentation layer, consists of the integration of several JavaScript libraries to create a graphical user interface that enables data and map visualization, data searching and manipulation, and data entry. The implementation of the user interface has included the development of some

specific components and the aggregation of several libraries: OpenLayers, GeoExt, ExtJS, jQuery and GeoNetwork. OpenLayers is a JavaScript library for displaying spatial data in a web browser, that supports OGC's standards (WMS, WFS-T, GML, SLD, KML) and other sources such as Google Maps, Virtual Earth, OSM, among others. GeoExt combines the map visualization capacity of OpenLayers with the user interface of ExtJS to allow the implementation of rich web mapping applications. ExtJS is a JavaScript library for the development of web-based desktop-like applications, that provides several graphical controls such as text fields, combo boxes, tab panels, grids, toolbars, windows, forms, among others. JQuery library was selected to parse XML documents when loading metadata files from catalog. Finally, GeoNetwork JavaScript library was used to develop the metadata searching and visualization interface of the BIO_SOS geoportal. The components of the presentation layer communicates with data access layer components through the use of web services, in particular Web Map Service (WMS) and Web Feature Service (WFS) for map visualization, as well as XML services to access several internal structures of GeoNetwork such as login services, users and groups services, and metadata services.

4.2 Data and metadata specification

4.2.1 Data specification and management

Spatial data in the metadata geoportal are meant to serve as a contextual basis for Partners (i.e. users) within the consortium, allowing them to visualize selected spatial datasets for each site and also, when available from public spatial databases, for European, Asian and South American regions. Within the metadata geoportal, spatial data will be gathered and visualized through a geographic interface that serves as a front face of the metadata geoportal, allowing the visualization, search and (in future developments of the platform) geoprocessing with simple spatial analysis tools.

In this context, the selection of spatial datasets to be represented within the metadata geoportal will follow a set of core criteria, namely: (i) the value/relevance of the dataset when compared with goals and objectives of the project (and of the geoportal within the project); (ii) the provision of contextual (essential) information; (iii) the identification of sites and of pre-identified critical information (e.g. BIO_SOS core datasets; see Section 5); and (iv) the inclusion of data produced in the context of the project life time (e.g. habitat maps). Considering these criteria, a set of datasets was selected (Table 4.1) and divided into two levels (i.e. Global and Site view), regarding the scale and geographic scope of the information. Although all available datasets could eventually be considered important and relevant for the BIO_SOS context, two moments/phases of implementation were defined in order to allow capacity tests on the server to be performed and possible problems to be corrected.

The selected datasets will appear on the metadata geoportal as viewable layers of geographic data, but no editing capabilities will be allowed. In this context, from the user's point of view, spatial data will only be used for visualization and interpretation, as well as for specific (limited) spatial queries subject to service providers' capacities (e.g. identification of all objects within a determined area). From the administrator's point of view, data management will be done using back-office tools and a specific procedure that will allow lodging all datasets within the metadata geoportal while considering all property and author recognition standards.

With the progression of the BIO_SOS projects and the gradual completion of its Tasks (particularly those related to WP5 and WP6), it is expected that new datasets and related data flows will become predominant. These new datasets and data flows should be included in the metadata geoportal in order to promote data interoperability, sharing and quality evaluation within the project. In the near future, the metadata geoportal will therefore consider the need to integrate these new spatial datasets produced within BIO_SOS, as well as the implementation of spatial

data quality management tools and protocols under a specific data lifecycle management framework (as previously discussed in D4.1 and described in Section 6 of this report).

Table 4.1 - Datasets that are available for contextual view in the BIO_SOS metadata geoportal.

Theme	Source	Format	Representation Scale	Web Service	Service Type
Global Environmental Stratification	Metzger et al., 2012	Raster (geotiff)	[global view]	-	-
Terrestrial Ecoregions	WWF	Vector (*.shp)	[global view]	-	-
GlobeCover (2005-2006)	ESA	Raster (geotiff)	[global view]	-	-
GlobeCover (2009)	ESA	Raster (geotiff)	[global view]	-	-
Global Land Cover Characterization	USGS	Raster (geotiff)	[global view]	-	-
WorldClim (several datasets)	WorldClim	Raster (geotiff)	[global view]	-	-
Ortoimagery	Bing/Google	-	[global view]	-	XYZ Grid/Google Maps API
World Countries		Vector (*.shp)	[global view]	-	-
World Regions		Vector (*.shp)	[global view]	-	-
Natura Network 2000	EEA	Vector (*.shp)	[site specific]	-	-
Biogeographic Regions	EEA	Vector (*.shp)	[global view]	-	-
European Environmental Stratification	Metzger et al., 2005	Vector (*.shp)	[global view]	-	-
Corine Land Cover (2006)	EEA	Vector (*.shp)	[site specific]	-	-
Corine Land Cover (2000)	EEA	Vector (*.shp)	[site specific]	-	-
Corine Land Cover (1990)	EEA	Vector (*.shp)	[site specific]	-	-
Roads (OpenStreetMap)	Bing	-	[global view]	-	WMS
Streets/Roads	OSM/Bing/Google	-	[global view]	-	XYZ Grid/Google Maps API
Soil Classification WRB	JRC	-	[site specific]	http://eusoils.jrc.ec.europa.eu/wrb/wms_WRB.asp?	WMS
FAO Soil Classification	JRC	-	[site specific]	http://eusoils.jrc.ec.europa.eu/wrb/wms_FAO.asp?	WMS
Test Site limits	Partners	Vector (*.shp)	[site specific]	-	-
Other conservation statutes	Partners	Vector (*.shp)	[site specific]	-	-
ASTER	JPL/NASA	Raster (geotiff)	[site specific]	-	-
Administrative Units of Europe	EuroGeographics	Vector (*.shp)	[site specific]	-	-
Local land cover maps	Partners	Vector (*.shp) or Raster (geotiff)	[site specific]	-	-
Hidrography	Partners	Vector (*.shp)	[site specific]	-	-
Habitat maps	Partners	Vector (*.shp) or Raster (geotiff)	[site specific]	-	-
Ortoimagery	BIO_SOS	Raster (geotiff)	[site specific]	-	-

4.2.2 Metadata management and catalogue services

WebGIS platforms, such as the one proposed to support the BIO_SOS metadata geoportal, can play a key role in promoting data sharing, with gains associated with resource use efficiency and with increased spatial data use and production. In fact, metadata include fundamental information about spatial datasets, such as purpose, quality, actuality and accuracy, but they also facilitate data sharing between systems (Rajabifard et al., 2009). In order to facilitate the access and use of metadata and spatial data, one of the key elements is the set of services accessible through the Internet that can be used by the geospatial community. Additionally, key factors for access to services in the Internet are interoperability and open standards, and, in the case of geospatial Web services, the Open Geospatial Consortium (OGC) standards have become the core standards to follow (Lopez-Pellicer et al., 2011). In the case of metadata information, catalogue services support discovery and access to geographic information resources, and thus they can be used to allow users to locate and access the appropriate data or services fitting their requirements.

In the context of the BIO_SOS project, it will be of high importance having the metadata platform set up with a catalogue services for the discovery and retrieval of data and metadata produced in the context of the project, in order to promote information sharing and to enable the interoperability and integration with internal and external

information systems. Therefore, the metadata platform includes the capacity to publish metadata using Catalogue Services for the Web (CSW) according to the OGC specifications (OGC, 2007), offering a standardized interface to publish metadata on geographic information, and providing mechanisms to search, access and maintain metadata information.

The implementation of CSW provides a way to publish and search collections of descriptive information for data, services, and related information objects (OGC, 2007), accomplished using a standard request-response model of the HTTP protocol between a client and a server using XML (figure 4.2). Therefore, the CSW provides a set of abstract service interfaces that support the discovery, access and maintenance of catalogues of geospatial information and can be used by client applications to manage and organize collections of metadata entries, to execute searches on metadata repository, and to receive results from the execution of queries.

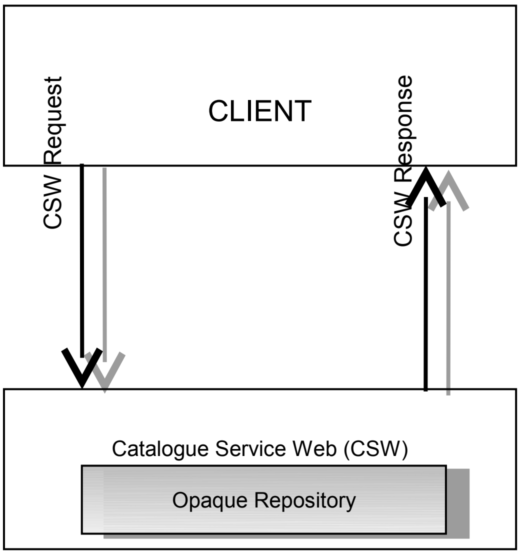


Figure 4.2 – Catalogue service web (from OGC, 2007)

Within the scope of the BIO_SOS WebGIS platform development, the catalogue services are available through the GeoNetwork open-source application, as it supports the publishing of metadata using the CSW protocol, compliant with version 2.0.2.

Among those defined in the CSW specifications, GeoNetwork implements the following five operations most relevant for the BIO_SOS platform (GeoNetwork opensource, 2011b): getCapabilities (to retrieve service metadata), DescribeRecord (to discover elements of the information model supported by the target catalogue service), GetRecordById (to retrieve the default representation of catalogue records using their identifier), GetRecords (to query the catalogue metadata records specifying a query in OCG Filter or CQL languages), and Transaction (for creating, modifying and deleting catalogue records). A more complete description of each operation can be found in OGC (2007).

Besides the CSW standard provided by BIO_SOS metadata platform, other OGC geographic Web Services standards, such as Web Map Service (WMS) for requesting georeferenced map images, Web Feature Service (WFS) for accessing raw vector data, and Web Coverage Service (WCS) for raw raster data access, can be easily implemented in the future, as the metadata platform includes software components with the capacity to implement these standards and to be configured as the platform evolve.

4.3 Internal and external availability and sharing of data and metadata

4.3.1 Evaluation of data availability

As stated before, the evaluation of data availability according to pre-determined standards and/or models or user requirements is one of the most important advantages of the use of a catalogue service, both from the scientific and from the application points of view. Regarding this context of implementation, the metadata geoportal is prepared to receive tools for the calculation of data availability metrics that allow users to determine which themes and datasets (and types of data) are available within the consortium, considering metadata actually loaded comparatively to metadata a priori defined for each site, process and/or application domain.

These tools will be based on geographic and thematic search engines in order to allow a quick view and establishment of the pre-requisites or selection criteria. Also, the system is prepared to return a summary of the more important results, namely: (i) the datasets available that meet the expected selection/evaluation criteria; and (ii) the level of completion of the selected criteria. These tools will be implemented according to the internal and external quality evaluation strategies defined and implemented in deliverable **D4.1** and in Section 5 of this report.

4.3.2 Implementation of data quality evaluation protocols within the metadata geoportal

Quality evaluation within BIO_SOS, and specifically within the metadata geoportal, is not exclusively related to the evaluation of available pre-existing datasets. In fact, quality evaluation (or quality guarantee) should have a broader scope including the quality protocols and the quality chain involved in maintaining a standardized fulfillment and availability of metadata within the metadata geoportal across the project and the consortium.

A quality chain was implemented considering the various domains in the metadata geoportal and the specific quality standards set for each domain. These domains include:

i) ***User identification and accreditation***

At this level, a user management structure was implemented based on user validation procedures that allow to certify different users to different access profiles. After a previous identification and acceptance by a specific user administrator, registered users will have to be accredited in the geoportal in order to be able to introduce new metadata or to edit pre-existing ones. Although this was technologically implemented in the platform, it will be necessary to establish specific data and user management policies. These policies will have to be as strict as required by the amount of data available and by the relevant property issues related to it. This is of high importance in order to convince users to engage in the usual utilization of the metadata geoportal and to implement knowledge networks supported by the communication strategies defined in the scope of the geoportal.

ii) ***Metadata fulfillment***

A metadata editor was implemented in order to standardize the fulfillment of the selected profile (see Section 5) and also to harmonize the language and the attributes considered for that task. The metadata editor was implemented having, whenever possible, a standard fulfillment based on pre-existing controlled lists derived from international standards.

The quality of the available metadata will depend upon a set of four steps: (i) identification of users; (ii) user accreditation for a specific user profile; (iii) fulfillment of standardized metadata according to the metadata editor; and (iv) validation, reviewing and publishing of the metadata by a reviewer.

4.3.3 Data availability and sharing in the context of the consortium

Although it wasn't a prime objective of this deliverable, possible solutions for data availability and data sharing in the context of the BIO_SOS consortium were considered. Ongoing efforts to implement these functionalities are dependent upon a possible data warehouse managed from the FTP available for the consortium, or other form of data server, and should consider the eventual data needs from EODHaM, that will likely indicate the necessity of data sharing protocols. Following the implementation of communication services to retrieve information on spatial data (metadata) from other catalogue services (outside the consortium), it will also be possible to identify and eventually access data present in other web based databases. In this sense, metadata sharing will be promoted in the geoportal by making it available for all partners within the consortium through user accreditation.

Within this development framework, data and metadata sharing in BIO_SOS will need to be implemented as two parallel segments: (i) data and metadata sharing among sites and partners (inside BIO_SOS) through the implementation of discovery services, indirect connections to an eventual FTP data warehouse inside the project, internal and external spatial data quality evaluation tools and protocols, and communication standards that facilitate the fulfilling of new metadata for spatial datasets produced within BIO_SOS, and particularly for those resulting from EODHaM; and (ii) data and metadata sharing among platforms and projects (outside BIO_SOS) through the implementation of catalogue services (CSW), discovery services, and visualization services (WMS and/or WFS). This last topic will allow not only to improve the communication inside and outside the project, but also to facilitate the dissemination and recognition of project results among a broader community of users (and uses).

At the present stage, metadata communication protocols are already implemented and the metadata geoportal is prepared to receive and to generate WMS and WFS for available data inside the consortium. Also, in order to implement data sharing dynamics, the metadata profile was prepared to integrate web links, or other form of web based connectors, to available data inside the consortium, promoting future data sharing among users.

4.4 The BIO_SOS geoportal interface

From a general perspective, the BIO_SOS metadata geoportal interface should provide a collaborative platform with the necessary tools to: (i) catalogue all relevant in situ and ancillary data across the consortium based on a common metadata profile, promoting the harmonization of dataset description; (ii) search and share metadata information among Partners; and finally (iii) the localization and visualization of reference and thematic datasets, in both raster and vector formats, organized according to the INSPIRE spatial data model specifications. According to these requirements, this section describes the BIO_SOS metadata geoportal interface and discusses issues related to: (i) software and geoportal implementation; (ii) user access and management; (iii) metadata input and management; and (iv) the geographic interface.

4.4.1 Software and geoportal implementation

The BIO_SOS metadata geoportal interface consists of three distinct parts accessible through a unique web page address: the geoviewer interface, the metadata management interface, and the user management interface. The geoviewer interface provides an interactive map that allows the visualization and visibility control of diverse reference and thematic cartographic data. The metadata management component focuses on the searching, access and editing of metadata entries. Finally, the user management component allows the definition of groups, users and their privileges regarding metadata entries access.

As previously stated, the metadata geoportal shall provide a simple and easy-to-use graphical user interface, therefore its implementation was based on an extensive use of several graphical libraries with the intent to create a rich web-based application. The implemented human-interaction solution, which has been developed through the

integration and customization of components of free and open source software and applications, is based on the combination of diverse graphical user interface (GUI) controls (or widgets) to enhance the efficiency and ease of use: (i) Ext JS 3.4 library, to build the interactive cross-platform desktop-style user interface of the BIO_SOS metadata geoportal, based on the combination of customizable visual building blocks; (ii) GeoExt, to create the interactive web-based mapping interface for viewing and styling geographic data; and (iii) GeoNetwork JavaScript library, to create the user interface which supports the discovery, visualization and access of metadata information.

One of the critical issues of the proposed platform is related with the collection of metadata based on the fulfillment of the predefined core metadata profile. In fact, the experimental tests performed on functionalities and customization of GeoNetwork online advanced metadata editor have shown that the editing interface was not as simple as needed for less-experienced metadata creators who are unfamiliar with complex metadata profiles such as ISO 19115. Therefore, in order to implement a user-friendly metadata collection process that copes with the difficulties of information gathering, a custom editor was developed to support the collection of the core metadata profile elements defined in the context of the BIO_SOS project.

The implemented approach was based on the development of a web form, which allows the contents of the metadata profile to be edited on-line, validated by automated procedures, and submitted to the catalog. In order to make the XML submission and loading process transparent to the user, a JavaScript extension has been developed to extend functionalities of Ext JS GUI controls so as to be able to generate and parse XML documents on the client side. These functionalities were achieved through the encapsulation of some application logics in the user interface controls of the Ext JS library. The implemented strategy consists in the integration of rules in the visual widgets definition of the forms layout that provides the logic for output generation, type validation and input parsing operations, in order to send requests and retrieve results using GeoNetwork XML metadata services.

The rules can be added to each GUI control through the definition of an object (identified by MD) that can contain the definition of several predefined properties to specify opening and closing tags to be used in the production and reading of XML elements, functions to format output values, to validate user inputs, and to process text elements from input XML files, as well as the definition of opening and closing tags to be used in the generation of row elements in grids. The generation and parsing of XML metadata profile uses a recursive procedure that processes GUI controls from an array with the definition of the ordered list of controls to be considered for writing and reading an XML document based on the International Standard ISO 19115:2003 and encoded according to the implementation schema ISO 19139:2007.

Based on the properties added to each Ext JS GUI control, the editor offers the capacity to generate metadata in XML format in order to insert or update GeoNetwork catalogue, and to receive an XML metadata document, process it to extract values from XML elements, and load their values into GUI controls. Additionally, some extensions to GUI controls have been developed to support metadata input, particularly a customizable GUI control extension of data grid panel in order to facilitate the adding, editing and removing of multiple entries of metadata elements.

The metadata model used in the implementation of the BIO_SOS metadata editor follows the INSPIRE metadata implementing rules, based on ISO 19115/19119 (EC, 2009), with a set of additional elements considered in the standard ISO 19115 (see Section 5). Figure 4.3 shows the elements for dataset identification included in the metadata editor, describing the properties of elements and their cardinalities following the notation presented in INSPIRE metadata implementing rules (EC, 2009, p.49), and including a reference to section number of Appendix I for each element. Regarding the encoding of metadata in XML format used in output and input operations, the guidelines established by the technical specification ISO 19139 (ISO, 2007) were followed in order to translate the metadata model into an XML format.

+ fileIdentifier[1] : CharacterString		
+ language[1] : LanguageCode	1.1	Metadata language
+ hierarchyLevel [1] : MD_ScopeCode		
+ contact [0..*] : CI_ResponsibleParty	1.2	Metadata point of contact
+ organisationName [1] : CharacterString	1.2.1	Organization name
+ contactInfo [1] : CI_Contact		
+ address [1] : CI_Address		
+ electronicMailAddress [0..*] : CharacterString	1.2.2	Email
+ role [1] : CI_RoleCode		
+ dateStamp [1] : Date	1.3	Metatadata date
+ referenceSystemInfo [0..*] : MD_ReferenceSystem	7.1	Reference System Identifier
+ referenceSystemIdentifier [1] : RS_Identifier		
+ code [1] : CharacterString	7.1.1	Code
+ codeSpace [1] : CharacterString	7.1.2	CodeSpace
+ identificationInfo[1] : MD_DataIdentification	2	Dataset Identification
+ citation [1] : CI_Citation		
+ title[1] : CharacterString	2.1	Resource title
+ date [1] : CI_Date		
+ date [1] : Date	2.2	Date of publication
+ dateType [1] : CI_DateTypeCode		"publication"
+ date [1] : CI_Date		
+ date [1] : Date	2.3	Date of last version
+ dateType [1] : CI_DateTypeCode		"revision"
+ date [1] : CI_Date		
+ date [1] : Date	2.4	Date of creation
+ dateType [1] : CI_DateTypeCode		"creation"
+ identifier [0..*] : RS_Identifier	2.5	Identifier - Unique Resource Identifier
+ code [1] : CharacterString	2.5.1	Code
+ codeSpace [1] : CharacterString	2.5.2	NameSpace
+ abstract [1] : CharacterString	2.6	Resource abstract
+ pointOfContact [0..*] : CI_ResponsibleParty	2.7	Responsible party
+ organisationName [1] : CharacterString	2.7.1	Organization name
+ contactInfo [1] : CI_Contact		
+ address [0..*] : CI_Address		
+ electronicMailAddress [1] : CharacterString	2.7.2	Organization email
+ role [1] : CI_RoleCode	2.7.3	Responsible party role
+ descriptiveKeywords [0..*] : MD_Keywords	4	Keywords
+ keyword [1] : CharacterString	4.1/4.2/4.3	Keyword from INSPIRE data themes
+ thesaurusName [] : CI_Citation	4.3.2	Originating controlled vocabulary
+ title [1] : CharacterString	4.3.2.1	Title
+ date [1] : CI_Date		
+ date [1] : Date	4.3.2.2	Reference date
+ dateType [1] : CI_DateTypeCode	4.3.2.3	Date type
+ resourceConstraints [0..*] : MD_Constraints	10	Contraint related to access and use
+ useLimitation [0..*] : CharacterString	10.1	Conditions applying to access and use
+ resourceConstraints [0..*] : MD_LegalConstraints		
+ accessConstraints [1] : MD_RestrictionCode	10.2	Limitations on public access
+ otherConstraints [1] : CharacterString		"OTHER_RESTRICTIONS"
+ spatialRepresentationType [0..*] : MD_SpatialRepresentationTypeCode	6.1	Spatial representation type
+ spatialResolution [0..*] : MD_Resolution	6.2	Spatial resolution
+ distance [0..1] : Distance	6.2.2/6.2.3	Resolution distance and unit measure
+ equivalentScale [0..1] : MD_RepresentativeFraction		
+ denominator [1] : Integer	6.2.1	Equivalent scale
+ language [0..*] : LanguageCode	2.8	Resource language
+ topicCategory [0..*] : MD_TopicCategoryCode	3.1	Topic category
+ extent [0..*] : EX_Extent	5	Extent
+ geographicElement [0..*] : EX_GeographicBoundingBox	5.1	Geographic bounding box
+ westBoundLongitude [1] : Decimal		
+ eastBoundLongitude [1] : Decimal		
+ southBoundLatitude [1] : Decimal		
+ northBoundLatitude [1] : Decimal		
+ temporalElement [0..*] : EX_TemporalExtent	5.2	Temporal extent
+ extent [1] : TM_Primitive		
+ Time [1] : TimePeriod		
+ beginPosition [1] : Date	5.2.1	Starting date
+ endPosition [1] : Date	5.2.2	Ending date
+ supplementalInformation [1] : CharacterString	2.9	Supplemental Information
+ distributionInfo [0..*] : MD_Distribution	8	Distribution Information
+ distributionFormat [1] : MD_Format		
+ name [1] : CharacterString	8.1	Format Name
+ version [1] : CharacterString	8.2	Format Version
+ transferOptions [0..*] : MD_DigitalTransferOptions		
+ onLine[0..*] : CI_OnlineResource		
+ linkage [1] : CharacterString	8.3	Online resource: linkage
+ dataQualityInfo [0..*] : DQ_DataQuality		
+ scope [1] : DQ_Scope		
+ level [1] : MD_ScopeCode		" DATASET "
+ report [0..*] : DQ_DomainConsistency		
+ result [0..*] : DQ_ConformanceResult	11	Conformity
+ specification [1] : CI_Citation	11.1	Specification
+ title [1] : CharacterString	11.1.1	Title
+ date [1] : CI_Date		
+ date [1]: Date	11.1.2	Date
+ dateType [1] : CI_DateTypeCode	11.1.3	DateType
+ explanation [1] : CharacterString		
+ pass [1] : Boolean		
+ lineage [1] : LI_Lineage	9	Quality
+ statement [1] : CharacterString	9.1	Lineage

Figure 4.3 – Metadata elements for datasets included in the BIO_SOS metadata profile.

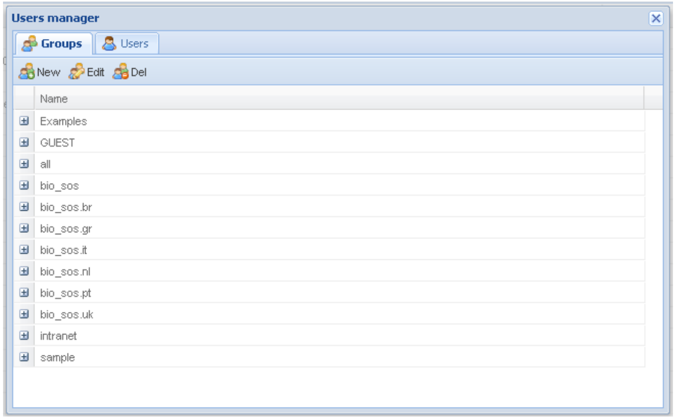
4.4.2 User access and management

The BIO_SOS metadata platform, regarding the user-access control, integrates components to ensure that access is only available to authorized users, thereby contributing to data quality and integrity. In this sense, the metadata geoportal requires users to log in using a username and password, in order to get access to information and functionalities according to their roles/profiles. Considering the predefined requirements, a user-access manager has been included in the metadata geoportal to allow the creation and management of users and associated privileges according to predefined administration levels, in order to allow to define permissions regarding the access, edition and publishing of metadata entries from the catalogue (see Section 3).

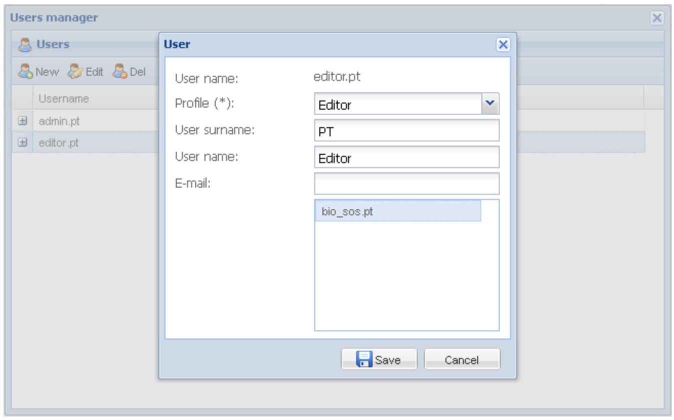
The user-access control module of the metadata geoportal has been developed based on user and group administration functionalities available on the GeoNetwork platform, as it complies with the requirements regarding users´ profile definition in the context of the project. The GeoNetwork opensource application (GeoNetwork opensource, 2011a) allows to define groups that correspond to logical units within the metadata platform, as well as users that belong to one or several groups and have a profile that defines the tasks the user can perform on the system and on specific metadata entries according to the groups to which he/she belongs. GeoNetwork defines five different user profiles as a hierarchical structure, where an upper profile inherits all privileges from a lower profile in the structure. The user profile hierarchy and associated privileges were illustrated in figure 3.1 (Section 3). Finally, access privileges can be set for each metadata entry, allowing to define the privileges for each group related to visibility of the metadata, possibility to download (meta)data, access to interactive maps, editing access (if the user has an editor user profile), and notification when a file is downloaded.

In the context of the BIO_SOS project, centered at this stage in the sharing and management of metadata information, the metadata platform was configured with a platform administrator, a group per set of Partners in charge of BIO_SOS sites in each country, and a global group to share metadata entries across the consortium. According to this structure, the BIO_SOS platform includes a graphical interface to define groups, users and access privileges for metadata entries, in accordance with each user’s profile. The platform administrator can create groups for new site Partners (Figure 4.4a) as well as the corresponding user administrator. For each site Partner group, the user administrator can create new users with reviewer, editor or registered privileges (Figure 4.4b). Within each Partner group, users with reviewing privileges can set the publication access level of metadata entries through the editor interface (Figure 4.4c), allowing to select between four publication levels: restricted to editor, visible to the users in the group of the editor, publication to all BIO_SOS partners group, or visible to all users that access the platform.

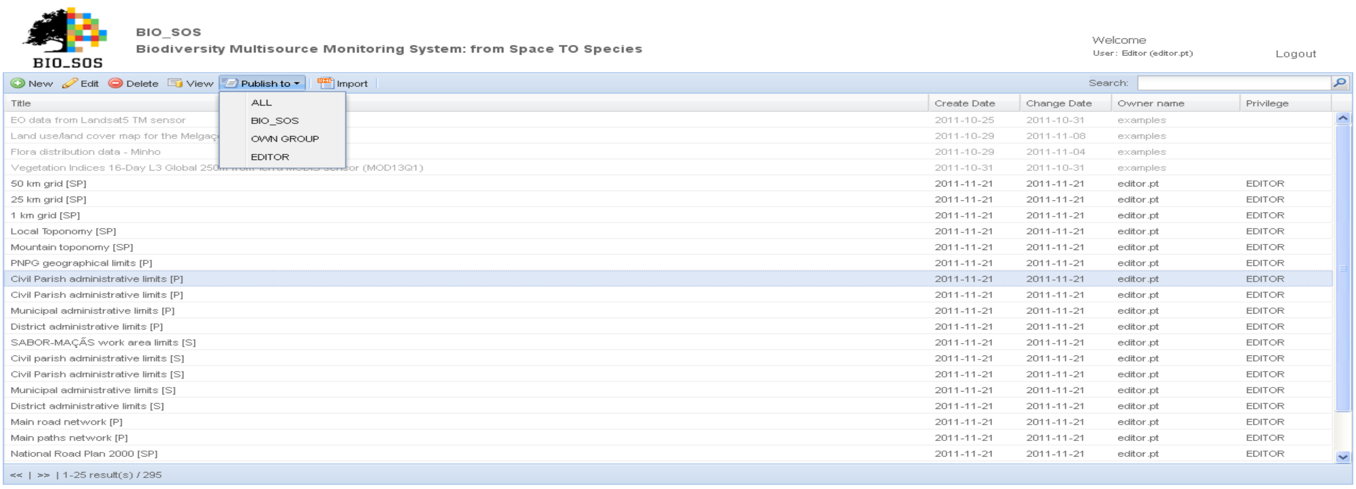
In order to allow the completion of the metadata profile for the collection of metadata on pre-existing datasets previously identified by each Partner (see deliverable **D4.1** and Section 5 of this report), the metadata platform was configured with a group and a user with a reviewer profile for each site Partners, besides the platform administrator. Methodologically, the general approach was to provide to each BIO_SOS site Partner an access to the metadata platform in order to fulfill the description of selected core datasets based on a complete metadata profile, using the metadata editor available through the metadata geoportal.



Group management



(b) User management



(c) Metadata publication management

Figure 4.4 – Interface for group (a), user (b) and metadata publication (c) management in the BIO_SOS metadata geoportal.

4.4.3 Metadata input and management

From a general perspective, this component of the metadata geoportal provides tools to query and retrieve information from the metadata catalogue, and to edit metadata information using a predefined metadata profile. In this sense, this component can be described as two independent modules: the geospatial metadata search page, and the metadata editor page. While the objective of the first module is the query and presentation of the results to the user, the aim of the second module is the creation, edition, modification, removal and publishing of metadata.

The geospatial metadata search page (Figure 4.5) provides functionalities that focus on the geospatial data discovery and retrieval, allowing to obtain complete information on the metadata that meet users' requirements. In this page, users can use different strategies designed to assist different types of queries for data searching and retrieval based on a simple or advanced search. The simple search option is a typical search that allows users to find metadata based on alphanumeric and/or geographic conditions, allowing to search entries containing one or more words within the entire record or limited to a geographic area of interest defined on an interactive map (Figure 4.6a). The advanced search provides to the user the possibility to set more search criteria on specific data elements to be used in finding metadata entries, such as search conditions on title, abstract, keywords, temporal extent,

scale, among others (Figure 4.6b). After the search execution, the list of matching metadata are presented in the results panel, where the user can click on list entries in order to view, print, download metadata contents or zoom to metadata geographic extents (Figure 4.7). The list of results is restricted to metadata records according to the authenticated user´s profile and privileges that have been set for each metadata entry.

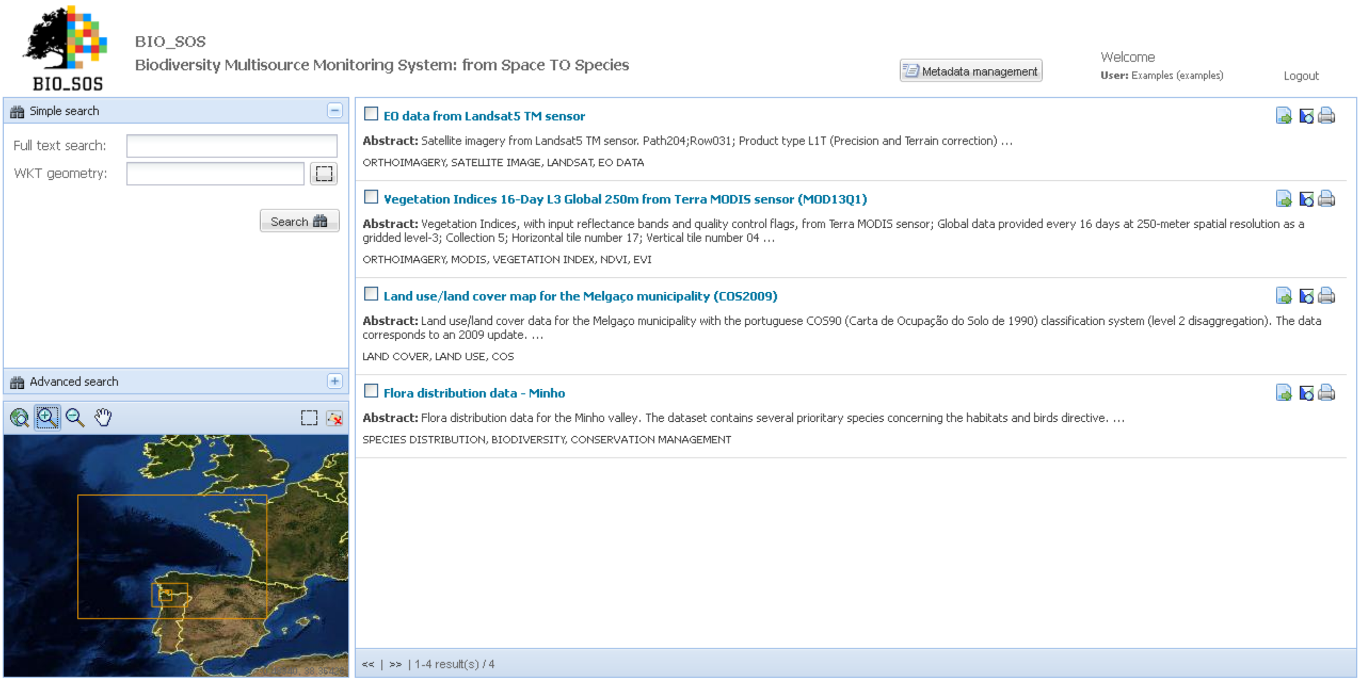
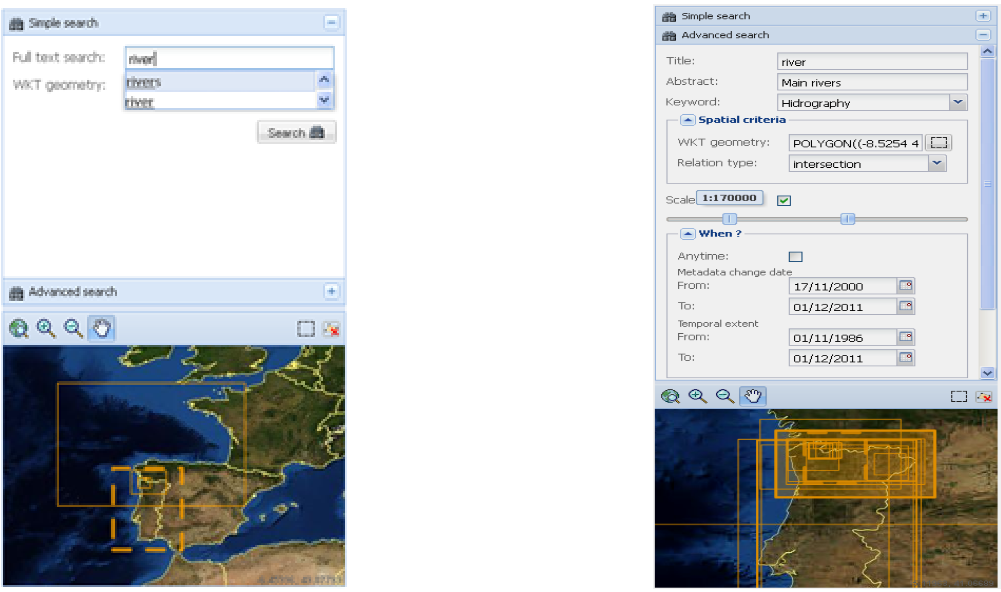


Figure 4.5 – Geospatial metadata search page in the BIO_SOS metadata geoportal.



(a) simple search (b) advanced search

Figure 4.6 – Simple (a) and advanced search (b) of metadata in the BIO_SOS metadata geoportal.

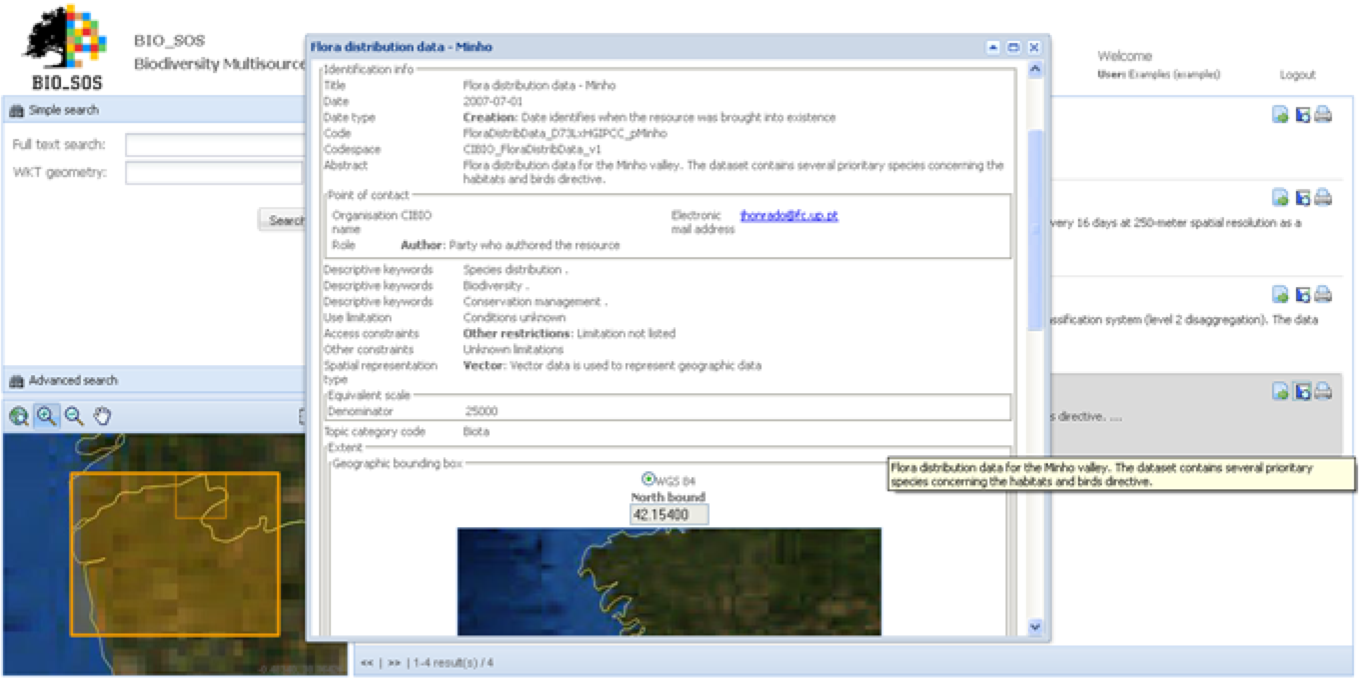


Figure 4.7 – Metadata contents view in the BIO_SOS metadata geoportal.

The metadata editor page is available for users with an editor profile (or superior profile) and provides tools to manage metadata entries in the GeoNetwork catalogue. The editor loads by default all the metadata records for which the logged user has reading privileges, and provides the possibility to restrict loaded records through the definition of an alphanumeric filter on the metadata title (Figure 4.8). The interaction with the metadata list can be achieved through the use of several tools available at the top toolbar, which allow the user to create a new metadata record, edit, remove or view an existing entry, and then publish metadata. The edition and removal operations are only available for metadata records for which the logged user has editing privileges, whereas publishing functionalities are only available to the reviewer user profile. In order to take advantage of all the work done during the metadata collection at the first stage of Task 4.1, and to create an initial state to support the updating/completing of information, all metadata collected by each Partner using the simplified metadata profile (see deliverable **D4.1**) were loaded into the metadata platform.

Considering that the fulfillment of metadata information can be a demanding and time consuming task when using a metadata profile based on a complex and large number of elements, e.g. from metadata standards such as ISO19115, the BIO_SOS metadata editor has been designed to assist metadata fulfillment for less-experienced users who are unfamiliar with metadata and standards such as XML, allowing to edit, validate and submit metadata contents to the catalogue.

Through the metadata editor page, the user has access to a metadata editor based on an on-line web form (Figure 4.9) created with the most appropriate graphical user interface component for the entry of each metadata element according to its data type. The form gives to the user the ability to edit metadata information by filling and selecting elements from the graphical interface using a variety of graphical input components such as text fields, lists, text areas, choices, combo boxes, date fields and grids, in order to create an easy-to-use editor. Whenever possible the form selection offers a list of predefined values to support fields completion (Figure 4.10a) and provides auxiliary windows e.g. for geographical extent definition (Figure 4.10b), and for completion of data with multiple values (Figure 4.10c). The editor also includes validation routines against standard rules and recommendations, particularly to comply with the metadata profile defined in the scope of the project (Figure 4.11).

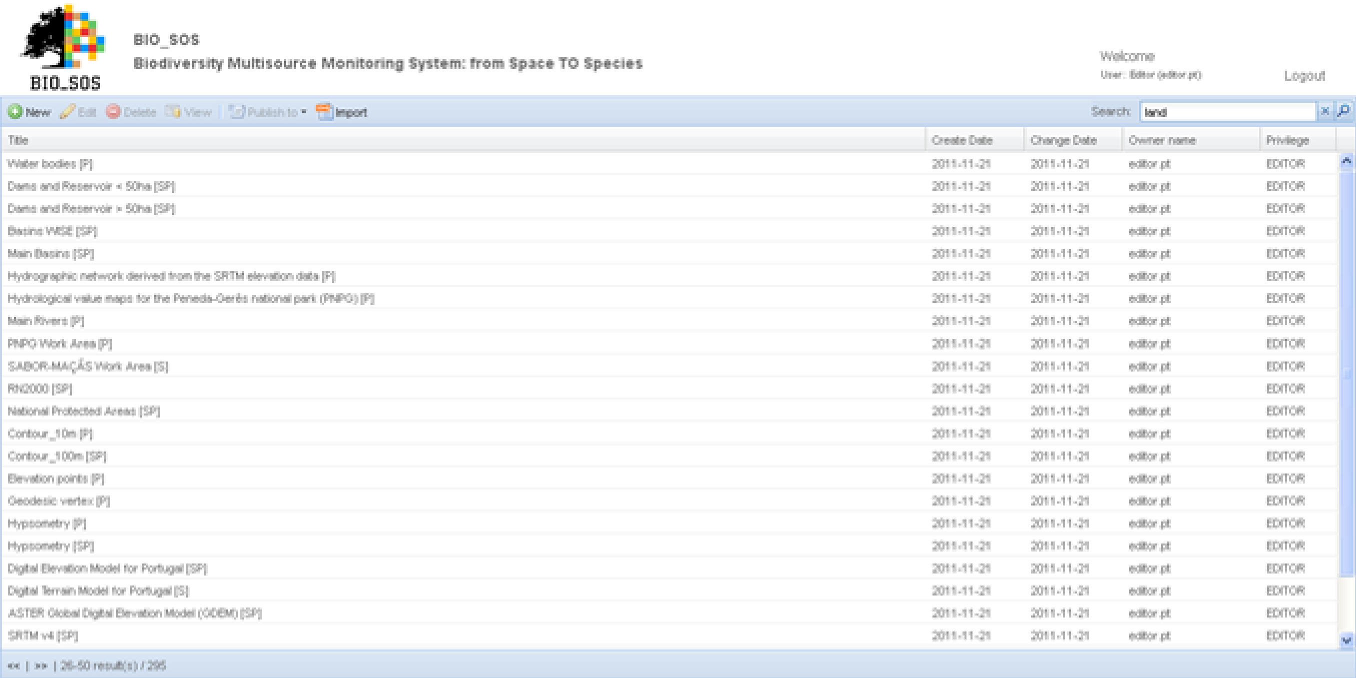


Figure 4.8 – Metadata editor page in the BIO_SOS metadata geoportal.

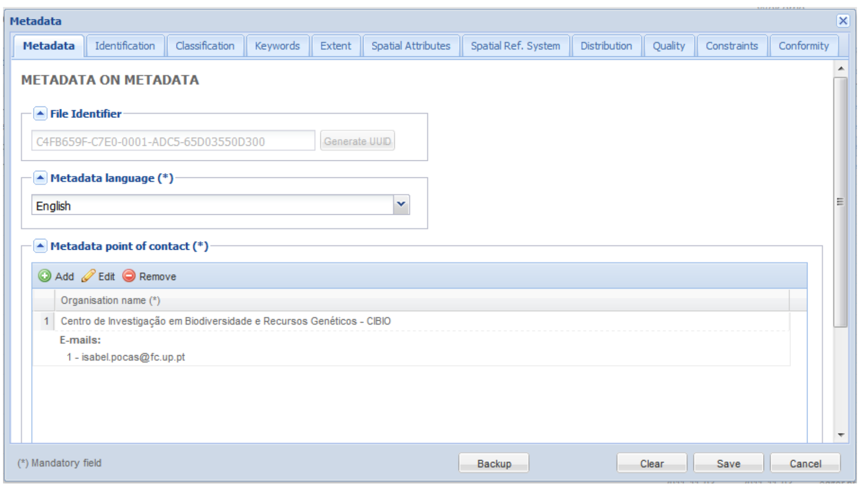
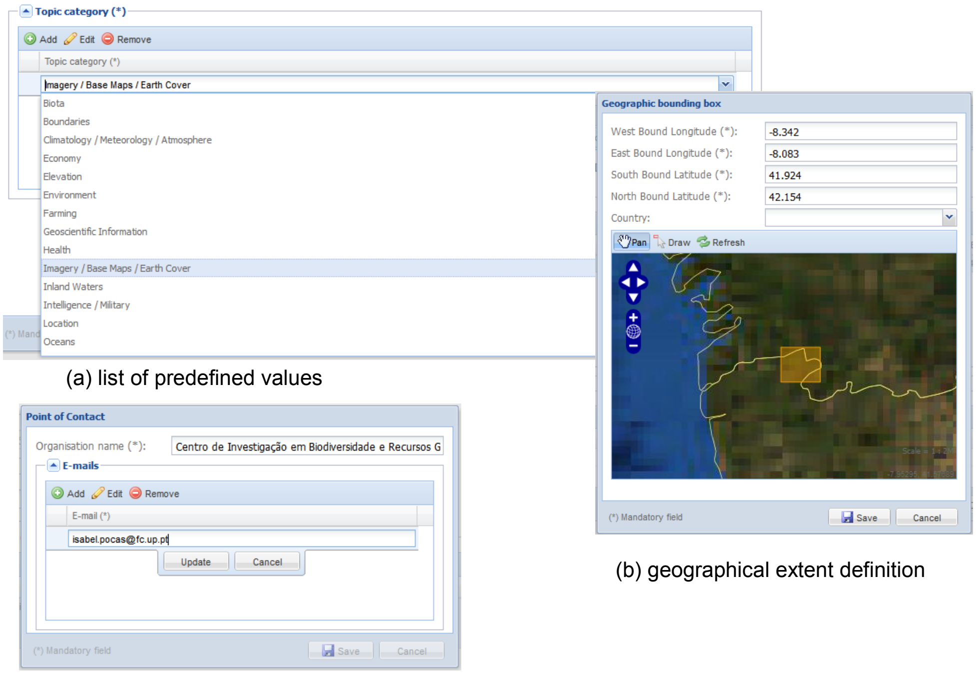


Figure 4.9 – Metadata on-line web form in the BIO_SOS metadata geoportal.



(a) list of predefined values

(b) geographical extent definition

(c) organization identification

Figure 4.10 – Graphical components to support metadata editing in the BIO_SOS metadata geoportal.

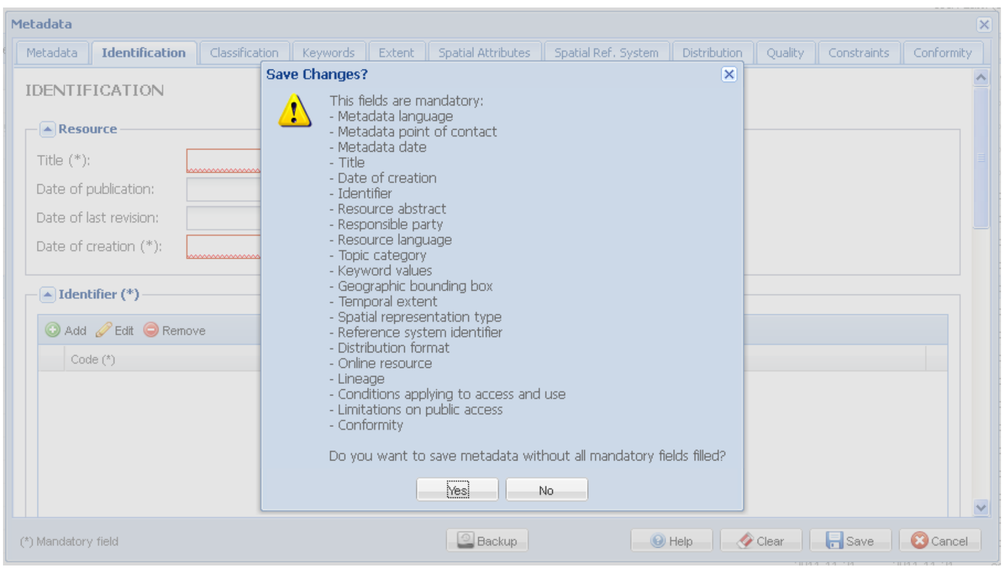


Figure 4.11 – Metadata validation in the BIO_SOS metadata geoportal.

Additionally, the editor includes an option to upload a new metadata entry to the metadata editor using an external document encoded in XML format and compliant with the metadata profile structure used in the project. After loading the data in the editor, the user can then submit the metadata entry to the catalogue.

4.4.4 Geographic interface

The BIO_SOS metadata geoportal includes a geographic interface that is the main entry interface of the system (Figure 4.12), divided in three parts: the login panel, the right panel with spatial layer visibility control, and the central panel with a dynamic map viewer.

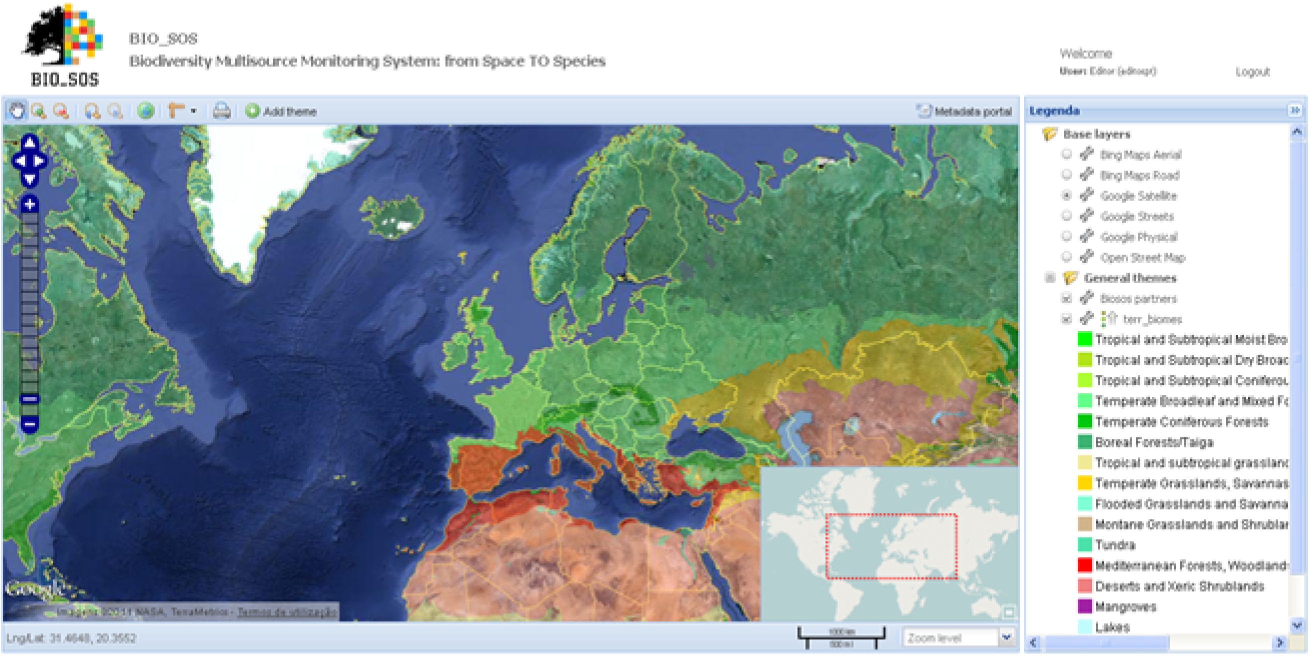


Figure 4.12 – Entry interface of BIO_SOS metadata geoportal

The central panel is based on an interactive map viewer with a set of general themes. The panel includes a toolbar with GIS common tools for map interaction, allowing navigating around the map, measuring distances and areas, and adding new layers available through external WMS service (Figure 4.13). The map also includes a locator map and a numerical combo box to control map scale. The right panel presents the available layers that can be activated/deactivated as well as their legends.

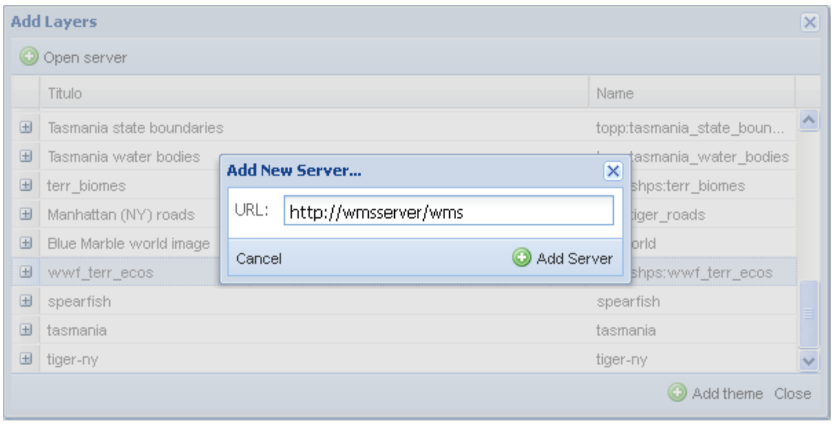


Figure 4.13 – Adding a WMS layer to geoportal

This main interface allows direct access to the other components of the BIO_SOS metadata geoportal for user management, spatial and thematic queries, visualization and metadata editing, according to the profile of the authenticated user.

5. EXTERNAL QUALITY EVALUATION OF PRE-EXISTING DATASETS

5.1 Introduction

As first described in **D4.1**, the external quality evaluation of core pre-existing datasets implemented for each site/Partner within the BIO_SOS consortium is based on the ontological approach defined by Vasseur et al. (2006). Such an approach allows the evaluation of similarities between user's needs or expectations, defined as the ontology of the problem, and the characteristics of the data (ontology of the product; Vasseur et al., 2006). In this methodological framework, the ontology of the problem and the ontology of the product are translated into quality matrices that represent the producer's perspective (intrinsic characteristics of the datasets) and the user's perspective (quality expectations).

An important issue regarding external quality evaluation is the dynamics of the process itself; i.e., it is not a static process because it may yield different results according to different users and/or different application contexts, but also because it can be seen as a continuous process towards the satisfaction of user's requirements. The next subsections of this report present the methodological framework for external quality evaluation in detail and the results of its application for different sites and considering several application contexts in the context of BIO-SOS goals and workflows.

5.2 Methodological framework

External quality is often defined as "fitness for use", since it corresponds to the degree of similarity between data characteristics and the user's implicit and/or explicit needs or expectations for a given application context and geographic area (Devillers et al., 2007). Therefore, for a given product or dataset external quality is not an absolute or invariable measure, but instead that same product or dataset can have different quality to different users and/or in distinct application contexts. This assumption is behind the methodological framework proposed in deliverable D4.1 for external quality evaluation (EQE) of pre-existing datasets (and also of newly acquired or produced datasets) within BIO_SOS; i.e., the proposed framework puts the user's context in the core of the approach and his/her requirements will determine the whole evaluation process.

EQE is a crucial issue for the determination of uncertainty, and therefore to avoid using data for which the potential negative impact on end-user's decisions is deemed unacceptable. Therefore this evaluation intends to help the user to evaluate the "fitness for use" of a dataset for a specific problem, providing a "user's confidence degree" that the inputs (datasets) used to derive products and outcomes to solve the problem will fulfil his/her expectations (or at least the minimum requirements). For the purpose of this deliverable and as part of work in Task 4.1, the datasets considered for EQE are the core pre-existing datasets identified by each Partner, which represent the most relevant datasets for a set of specified application contexts.

Given these considerations, the following sub-sections will present: (i) the complete metadata profile used to describe pre-existing datasets; (ii) the main application contexts within the BIO_SOS project and the guidelines for selection of core datasets; and (iii) the details of the methodological framework for EQE.

5.2.1 The complete metadata profile

Metadata are data about the content, quality, condition, and other characteristics of data (Federal Geographic Data Committee, 2000). Therefore, metadata describe the characteristics of a resource, e.g. geospatial data. The increase in the amount of data developed by the geospatial community in the last decades generated the need for using standard tags for the description and characterization of datasets. The international standard ISO 19115: 2005 ("Geographic Information – Metadata") provides the information required for describing geographic information, allowing to: (i) determine the availability of a set of geospatial data, (ii) determine the fitness for use of a set of geospatial data for specific uses, (iii) determine the means to access the set of geospatial data, and (4) transfer the set of geospatial data successfully. Therefore, it enhances data understandability and data discovery functionalities and allows improving the cooperation and interoperability among organizations.

The inventory of pre-existing spatial datasets implemented in the first stage of Task 4.1 (Deliverable **D4.1**) was based on the characterization of databases for each test site through a simplified metadata profile. The information then collected, and now available in the metadata geoportal integrated in the new BIO_SOS WebGIS collaborative platform, provided a brief description of the available datasets for each site, allowing geoportal users to consult and discover information on topic category, spatial resolution, temporal extent, date of publication, geographic bounding box, file type, author, property, and spatial reference system.

In the second stage of Task 4.1, a complete metadata profile was defined to describe the core datasets identified for each site by BIO_SOS partners (listed in section 5.2.2). The complete metadata profile follows the concepts, themes and framework of the INSPIRE metadata regulation (i.e., the elements are compliant with INSPIRE metadata specifications) and it also includes a few supplementary elements, all in agreement with rules established by the international standards ISO 19115:2005 and ISO 19139:2007. This complete profile (here proposed as a "BIO_SOS standard metadata profile") identifies the metadata elements required to accurately describe the datasets used for the several WPs/Tasks of the BIO_SOS project and intends to support the documentation, search and use of datasets within the consortium and the sharing of information among Partners. This complete metadata profile allows and facilitates data cataloguing and management, not only of the pre-existing datasets (with a focus on core datasets, for the purpose of D4.5) but also of the new datasets to be acquired or produced in the project.

Figure 5.1 summarizes the metadata elements considered in the profile used in this evaluation. It provides information about the identification, the extent, the spatial and temporal schema, the spatial reference, the geographic location, the constraints related to access and use, the quality, and the distribution of different types of datasets. A detailed description of the metadata elements is provided in Appendix I of this report, based on European Commission Joint Research Centre (EC, 2009) and ISO 19115:2003.

The complete metadata profile was filled in by each Partner for his/her pre-selected core datasets (see sub-section 5.2.2), following the guidelines in Appendix I and using a metadata editor integrated in the new BIO_SOS WebGIS collaborative platform. The metadata information on core datasets was then used for the EQE presented in Section 5.3.

Level 1	***Level 2***	***Level 3***	***Level 4***
Metadata on metadata	Metadata language		
	Metadata point of contact	Organization name	
		Organization email	
	Metadata date		
Dataset identification	Resource title		
	Date of publication		
	Date of last version		
	Date of creation		
	Identifier - Unique Resource Identifier	Code	
		Namespace	
	Resource abstract		
	Responsible party	Organization name	
		Organization email	
		Responsible party role	
	Resource language		
	Supplemental Information		
Classification of spatial data	Topic category		
Keyword	Keyword from INSPIRE data themes		
	Keyword from online repositories		
	Free keyword	keyword value	
		Originating controlled vocabulary	Title
			Reference date
			Date type
Extent	Geographic bounding box		
	Temporal extent	Starting date	
		Ending date	
Spatial attributes	Spatial representation type		
	Spatial resolution	Equivalent scale	
		Resolution distance	
		Unit of measure	
Spatial reference system information	Reference System Identifier	Code	
		CodeSpace	
Distribution information	Format Name		
	Format Version		
	Online resource: linkage		
Quality	Lineage		
Constraints related to access and use	Conditions applying to access and use		
	Limitations on public access		
Conformity	Specifications	Specification Title	
		Date	
		Date type	
		Degree	

Figure 5.1 – Metadata elements considered in the complete metadata profile.

5.2.2 Definition of application contexts and selection of core datasets

As previously mentioned, the methodological approach for the EQE of datasets is “application context-specific”. Therefore, one of the first steps of the EQE process should be the identification of the application context(s) for which the evaluation is being performed. Within the BIO_SOS project, three main application contexts can be considered:

i) Production of habitat maps within the specified quality requirements for BIO_SOS (as defined by end-users and described in D4.3 and D5.1);

Pre-existing (and newly acquired) datasets will be used as input for EODHaM 2nd and 3rd stages, as described in the Service of Chain of deliverable D3.1 (e.g., pre-existing land cover maps, air photos, satellite imagery, transportation network information, …).

ii) Sampling design for the collection of new on site campaigns for: (a) calibration or validation datasets for EODHaM habitat maps; (b) collection of new data on biodiversity; (c) collection of new data on pressures;

Pre-existing (and newly acquired or produced) datasets can be used to support the sampling design at several levels (as described in deliverable D4.3): (i) for the collection of new on-site data of vegetation, fauna, soil (e.g., pre-existing land cover (LC) time series, topographic maps, …); (ii) for the collection of validation data on land cover/land use, GHC and Annex I habitat types (e.g., pre-existing LC maps and GHC maps,…); (iii) for the recording and mapping of pressures/threats (e.g. local collection of census data to identify areas submitted to specific pressures).

iii) Evaluation of the effects of specific pressures on habitats and biodiversity.

Pre-existing (and newly acquired or produced) datasets can be used to support the evaluation of the effects of specific pressures on habitats and biodiversity (e.g., pre-existing and newly LC maps, produced GHC maps, …) through modelling tools: (i) to provide data for ecosystem state and functions assessment, as required to assess soil/vegetation interactions, soil physical degradation, stoniness; (ii) to provide data on fine spatiotemporal threats such as soil degradation, land use and land abandonment, as required to demonstrate the adequacy of the BIO_SOS approach across a range of pressures in and around Natura 2000 areas; (iii) to create a dataset on landscape pattern at multiple scales (e.g., local, landscape) in order to explore their potential for predicting both local biodiversity attributes (e.g., species richness, abundance, diversity) according to ecological theory, and ongoing pressures and threats, as required for indicator estimation.

The core datasets to be used in each application context can be identified from the pre-existing datasets available for each site. Starbuck (2001) defines “core dataset” as the ‘core’ or ‘minimum’ set of spatial information needed to support national and subregional environmental assessment, decision making, and environmental reporting. Nerbert (2004) also defines “core data” as a set of geographic information that is necessary for optimal use of most GIS applications. Gyamfi-Aidoo et al. (2006) define “fundamental geospatial datasets” as the “minimum primary set of data that cannot be derived from other datasets, and that are required to spatially represent phenomena, objects, or themes (…) at local, national, sub-regional and/or regional scales”. Among the many criteria and guidelines for identifying fundamental geospatial datasets (adapted from Gyamfi-Aidoo et al. 2006; RDM Working Group 2002), these should:

i) Include a complete coverage over the area of interest;

ii) Include sufficient detail (considering the requirements of different application contexts);

iii) Have acceptable standards and validation processes that ensure consistency, reliability, quality, continuity and accuracy (since these datasets are those upon which other thematic datasets and products will be developed);

iv) Provide a context to allow other potential users to better understand the information that is being presented;

v) Be collected by official agencies;

vi) Enable the merging of data from various sources, i.e., integration of data from heterogeneous sources in a network environment, which requires a neutral format and an underlying conceptual model; and

vii) Allow that a diversity of users from different sectors may derive significant benefit from their use.

Considering these definitions, and as previously described in **D4.1**, the identification of core/fundamental datasets within the collection of pre-existing datasets from each BIO_SOS site should support decisions on a wide range of application contexts within the scope of the project (i.e. high-quality and high-resolution habitat and biodiversity monitoring). Thus, for the purpose of EQE within BIO_SOS, the operational concept of core/fundamental datasets can be described as the most relevant datasets to support (i) the assessment of the process(es) in study by each Partner at each site, and (ii) the overarching goals of BIO_SOS. Therefore, the selection of core datasets should identify the fundamental datasets considering the above mentioned application contexts.

Considering the diversity of sites and their characteristics, as well as the diversity of processes/pressures being studied, the core geospatial datasets need to be identified within appropriate user-defined frameworks, and thus contribution from site Partners was required for the identification of core datasets for each site. The core datasets identified by each Partner according to specific application contexts are presented in Appendix II, and Figure 5.2 summarizes the number of core datasets per application context and site/country.

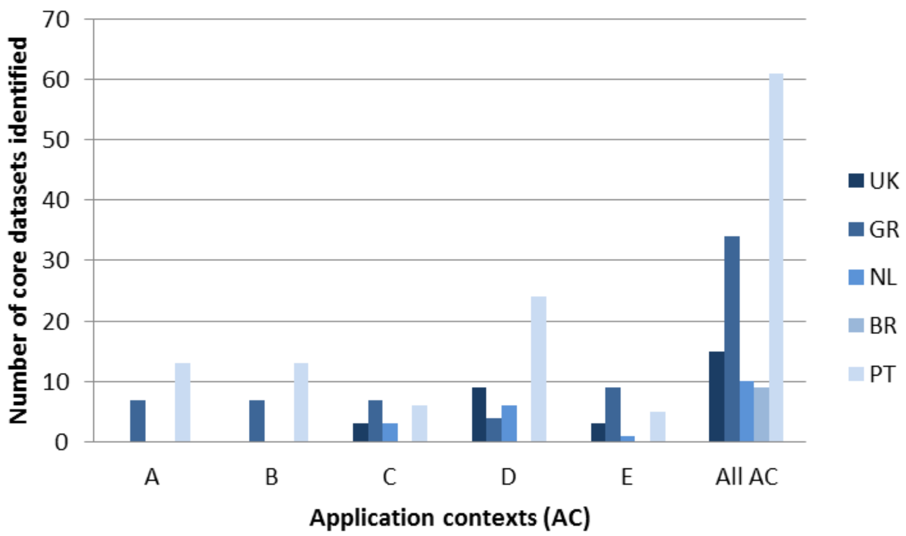


Figure 5.2 – Number of core datasets identified per application context and Site/Partner. Application contexts: A: Sampling design for the collection of new on site campaigns to collect new data on pressures; B: Sampling design for the collection of new on site campaigns to collect new data on biodiversity; C: Sampling design for the collection of new on site campaigns for calibration or validation datasets of EODHaM habitat maps; D: Production of habitat maps within the specified quality requirements for BIO_SOS; E: Evaluation of the effects of specific pressures on habitats and biodiversity. No information on Italian sites was available for this EQE.

5.2.3 External quality assessment

5.2.3.1 Introduction and definitions

The methodological framework implemented for EQE is based on an end-user, objective-oriented and data-centric perspective, meaning that the external quality of datasets will be assessed for each application context by measuring the overall matching between the characteristics of the data as detailed by metadata and the characteristics of the data as required by users.

The current methodological framework places the user's context, experience, knowledge and requirements at the core of the process, as his/her requirements will determine all the evaluation process. Nevertheless, it also strongly depends on the metadata actually residing on catalogue databases. This occurs because EQE as a process is not totally isolated from internal quality evaluation (derived from metadata; see **D4.1**), rather it actually requires information describing internal quality residing on appropriate database engines. In this sense, an adequate metadata profile, the accurate fulfillment of its elements, and a suitable definition of expected quality parameters are paramount to allow a correct comparison between internal (derived from metadata) and expected (required by the user) data quality indicators.

According to this framework, we hereby define:

- *Internal quality*: the internal characteristics or attributes of a given product (e.g.: a dataset, in this specific case) as described at the producer level and usually detailed in metadata;
- *Expected quality*: a set of user-defined values for several quality indicators that specifically describe the user's requirements; expected quality values should translate a pragmatic, self-contained and critical set of parameters allowing the assessment of data fitness-for-use in a certain application context;
- *External quality*: the level of similarity between the characteristics of a product (or dataset in this case) and the user's needs or expectations, in a given application context.

In the context of the proposed methodology for evaluating external data quality ("fitness for use"), we anticipate that this can be used for two distinct and complementary objectives:

i) External quality evaluation as an assessment procedure or diagnostic tool; or

ii) External quality evaluation as a tool for searching and ranking datasets according to external quality concepts.

For better understanding the purposes and the differences between these two options we will briefly characterize them in the following paragraphs.

i) External quality evaluation as an assessment procedure or diagnostic tool

EQE can be used to perform an assessment or diagnosis of pre-selected datasets (core datasets). Specifically, this can be used to:

- Evaluate the feasibility of a given analysis in terms of data availability and adequacy (fitness-for-purpose);
- Detail which datasets may hinder the ability to perform certain analyses and for those specify which data characteristics/features are considered problematic;
- Perform a progressive, iterative and continuous quality control and assurance over data products (e.g. those generated by the BIO_SOS consortium).

This option also allows a fine-grain definition of expected quality values compared to a data searching routine. While in the case of EQE use as a tool for searching and ranking datasets we are mainly concerned in finding and ranking datasets that correspond to a generic parameterization of quality, in the case of EQE use as an assessment procedure or diagnostic tool, quality indicators are defined at dataset level to allow a finer degree of control.

At a first stage, after selecting this analytical option, the user should be prompted to select which core datasets he/she intends to analyze given an application context. The selection of core datasets should be carried out based on the full list of datasets residing in the metadata catalog. Also, a pre-selection of datasets through an *a priori* filtering procedure can be implemented to present the user a reduced list more related to his/her intentions. Core dataset selection is not regarded as a previous external quality assessment, and it is performed intuitively by the user and based on his/her *a priori* knowledge. Instead, it should be considered as a step to limit the number of datasets to those that: (i) the user considers potentially useful to perform a given task; and (ii) quality control procedures will be applied upon. In this sense, core dataset selection can be based on a brief identification and description of data (e.g. through simplified metadata profiles; see **D4.1**).

This option should also allow users to store/save core datasets they have selected, expected quality values and summary diagnostics for further reference. This feature is potentially useful for iterative quality control over BIO_SOS data products.

ii) Data search, querying and ranking using EQE concepts

Commonly, spatial data searching routines do not produce a diagnosis detailing which datasets are fit or unfit and to what extent they are suitable or unsuitable for a given purpose. Usually an alphabetically sorted list is presented showing which data items strictly match the search terms, thus hindering the user to see other items that may be close to his preferences. Note that this is only possible in cases where the number of searchable elements is low and therefore the presentation of results is not very extensive.

As previously mentioned in deliverable **D4.1**, EQE and data searching mechanisms share many similarities and can be integrated within a single framework. This can be done by incorporating EQE concepts in data searching routines and especially by considering EQE as a special form of ranking and data criticality evaluation. Defined in such a flexible way, this EQE methodological framework will not only segment datasets into 3 classes – fit, partially fit, and unfit –, it will also provide a numerical fitting value (in percentage) and a summary describing where problems (may) have occurred. We argue that, by using this external qualitative description, a richer and more informative summary is presented to the user which enhances his/her capability to make decisions on whether to use or not some dataset. The search summary will use EQE concepts as a ranking algorithm similar to the way search engines present prioritized and ordered search results to web users.

In this case also an a priori filtering mechanism may be used to limit data searching procedures to some topic categories or keywords found in some metadata fields, e.g. abstract, free keyword, lineage, etc.

At this stage of Task 4.1, since the metadata editor and search tools were being developed as part of the BIO_SOS WebGIS collaborative platform, EQE was implemented outside the platform in a pilot-computational implementation in R language. This computational implementation was supported by specific R packages, e.g. rgdal, rgeos, sp and gdata. Nevertheless, the EQE methodology is already structured for integration in the WebGIS platform and will be integrated soon, interconnected with the metadata editor and with search tools. Therefore the technical specifications for the methodological framework detailed in the following sub-section (5.2.3.1) will already include references to EQE implementation in the WebGIS platform.

5.2.3.2 *Technical specifications*

In this section, the steps for the EQE at application level are presented. Figure 5.3 presents an algorithm overview of the series of steps that must be carried out to perform an EQE.

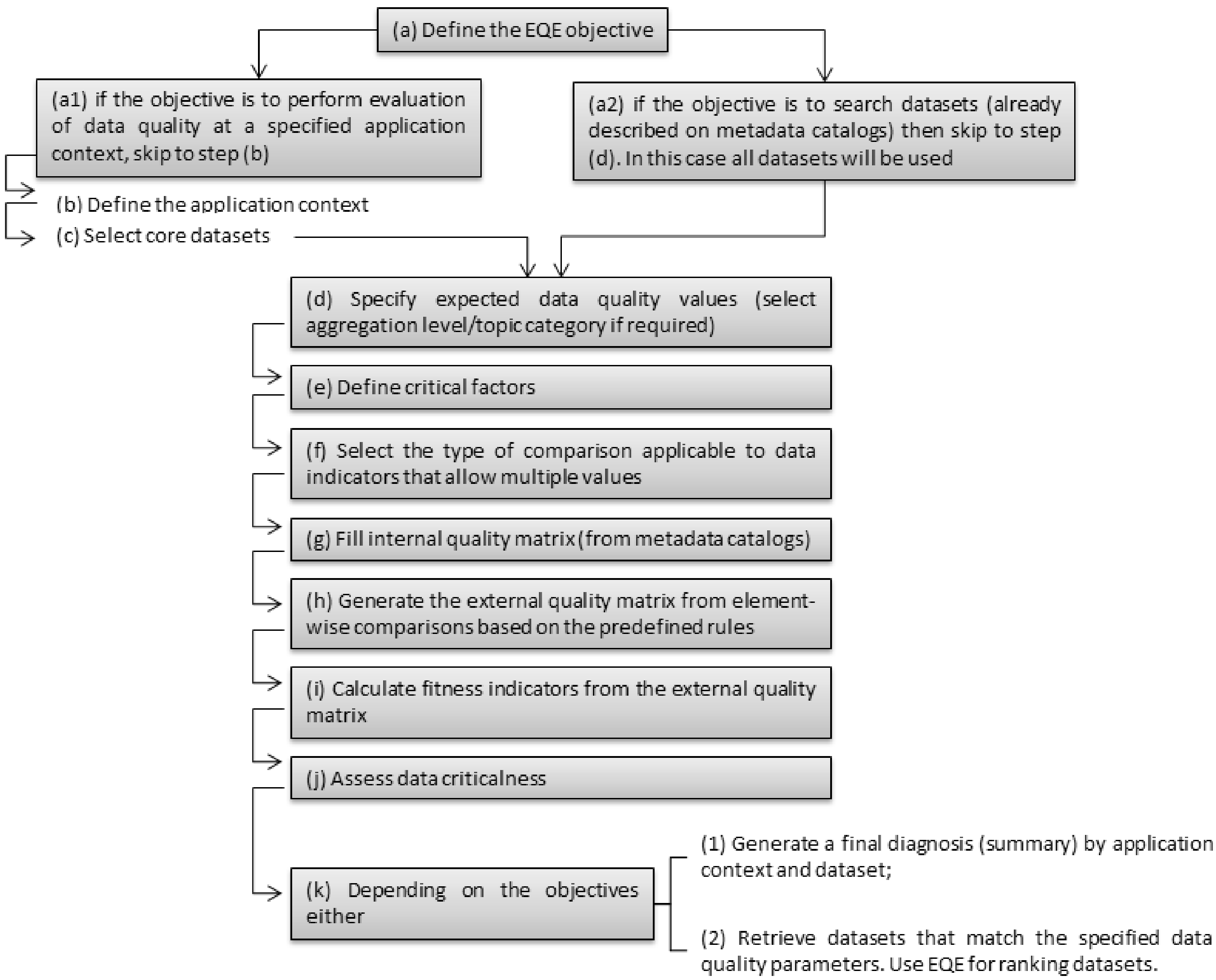


Figure 5.3 – General workflow for the EQE framework application in BIO_SOS.

- *Step a – Define the EQE objective*

This step consists of selecting which type of evaluation the user wants to perform. This decision strongly impacts the way quality assessments are performed as well as the graphical interface that interacts with the user. At this stage we have considered two mutually exclusive objectives:

(a1) external quality assessment using a set of core datasets, or,

(a2) searching and ranking datasets according to EQE concepts.

For the purpose of this deliverable only option (a1) was considered.

If option a1 is selected then the user will be prompted to:

- *Step b – Define the application context*

This consists of a brief description that states for what purpose the datasets will be used.

- *Step c – Select core datasets*

This step is used to define a subset of the dataset pool that will be specifically targeted for EQE. A form of data filtering and ordering should be used to display all available datasets in the catalogue in a manner that eases the selection process (see A1-I1 interface, in Appendix III).

The following steps are common both for options a1 and a2.

- *Step d – Specify expected data quality values*

- *Step e – Define critical factors*

- *Step f – Select the type of comparison applicable to quality indicators that allow multiple values*

These three steps consist in the definition of expected quality values (d), critical factors (e) and the type of comparison used if multiple values are allowed and defined (f). These steps run in parallel for each quality indicator. If option a1 was selected then steps d, e and f are declared for each dataset otherwise, if option a2 was chosen, then expected quality values are defined generically as search and ranking terms (See interfaces A1-I2, A1-I2a and A2-I1, in Appendix III).

- *Step g – Fill internal quality matrix*

After submitting expected quality values into analysis routines on the EQE application, the internal quality matrices are filled in according to the metadata database catalog.

- *Step h – Generate the external quality matrix from element-wise comparisons based on the predefined rules*

This step is used to perform the comparison between internal and expected quality values, and it is centered on a pre-defined rule-based system which controls how the pairwise comparisons are carried out. This step is responsible for producing an external quality matrix which in turn will allow the calculation of fitness values at the dataset level ("fitness for use") and/or at the application context level ("fitness for purpose"). This matrix will also enable the evaluation of critical factors.

Quality matrices are central to the proposed evaluation framework and they are formed by $i = \{1,2,\dots,m\}$ rows, with m being the number of datasets, and $j = \{1,2,\dots,n\}$ columns, with n representing the number of quality indicators. The matrices of internal ($A = [a_{i,j}]_{m\times n}$) and expected ($B = [b_{i,j}]_{m\times n}$) data quality must be established within the same geo-semantic reference frame, thus allowing a comparison between A and B.

Comparisons between expected and internal quality values are performed element-by-element for each listed dataset (row-wise). The comparison between A and B generates a final external quality binary matrix ($C = [c_{i,j}]_{m\times n}$) where each element is within the Boolean domain $c_{i,j} = \{0,1\}$; values equal to 1 correspond to conformities, and 0 (zeros) to non-conformities. In general, a higher degree of overlap between these matrices is indicative that the spatial data element fulfills or covers better the end-user's requirements specified for a given objective or application

context. This approach has an analogy with the perspective of gradual influence of the dataset quality on the user's decision about the suitability of data for a specific application context, as described by Frank et al. (2004): "the higher the quality, the better the decision".

- *Step i – Calculate fitness indicators from the external quality matrix*

From the external quality matrix, it is possible to calculate, for a given dataset i, its external quality (fitness for use) for a given application context k (i.e. Q_i; see Eq. 1), as well as its overall external quality for the specified application context (i.e. Q_k; see Eq. 2). Note that equation 2 is only used when the user selects core datasets (i.e. option a1 in Figure 5.3).

$$Q_i = \left(\frac{1}{n}\sum_{j=1}^{n} c_{i,j}\right) \times 100 \qquad \text{(Eq. 1)}$$

$$Q_k = \left(\frac{1}{m \cdot n}\sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\right) \times 100 = \left(\frac{1}{m}\sum_{i=1}^{m} Q_i\right) \times 100 \qquad \text{(Eq. 2)}$$

- *Step j – Assess data criticalness*

Critical factors (CF) are a central concept in the evaluation of external quality as defined in the proposed methodological framework. Critical factors are declared by users (as a complementary step when defining expected quality values) and are used to flag important quality indicators that if non-conformant will make the dataset automatically unfit for use.

The evaluation of critical factors follows the generation of the external quality matrix and is performed at the dataset level and for each quality indicator using the decision tree presented in Figure 5.4. When a quality indicator is considered a CF and a non-conformant result is obtained, this (or these) case(s) must be flagged and presented to the user by the application (for a summary, see interfaces A1-I3 and A2-I2 in Appendix III). Quality indicators that do not meet expected quality patterns specified by the user are vital pieces of diagnostic information and will help decision making processes.

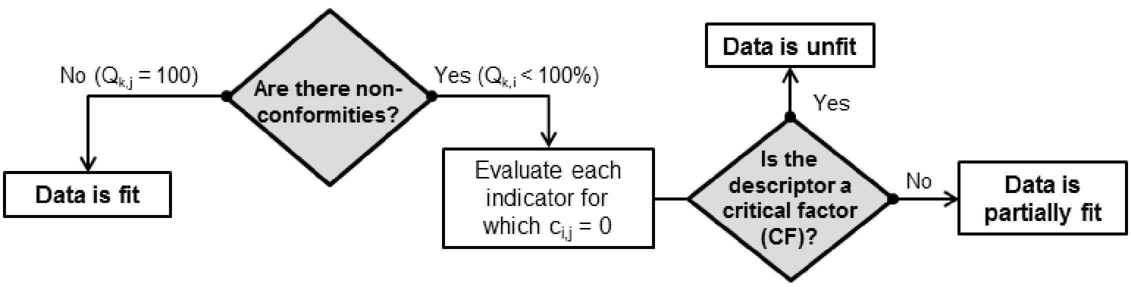


Figure 5.4 - Decision tree used to assess critical factors

The assessment of critical factors allows the classification of datasets into three broad categories:

iii) Fit (no non-conformities detected, all expected values match internal quality patterns);

iv) Partially fit (at least one non-conformity detected which is not a CF);

v) Unfit (at least one non-conformity detected which is a CF);

For the implementation of the EQE methodology in the WebGIS platform, a simplification of this step will be used, creating a vector of *n* columns (corresponding to the number of quality indicators) with ele-

)

ments, cf_j, represented by Boolean values {1,0} (1 = critical factor; 0= non-critical factor). The vector will be used to calculate a data criticalness factor (CF_i) for each dataset according to equation 3.

$$CF_i = \sum_{j=1}^{n}\left((1 - C_{i,j}) \times cf_j\right) \quad \text{(Eq. 3)}$$

Based on CF_i, the datasets will be classified as:

- Unfit, if CFi > 0;
- Partially fit, if CFi = 0 and Qi < 100%;
- Fit, if Qi = 100% (and CFi = 0).

Step k – Summary results

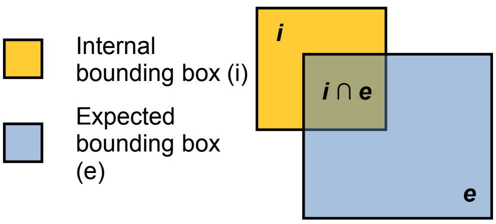
Depending on the objective set by the user:

- If option a1 was selected, then a final diagnosis or summary by application context and dataset will be generated;
- If option a2 was selected, the datasets that matched the specified data quality indicators will be retrieved, and external quality diagnosis will be used for ranking and organizing datasets.

Comparison terms between internal and expected quality matrices - Spatial or temporal intervals

When comparisons between internal and expected quality values are held for spatial bounding-boxes (spatial extent/2D) or temporal intervals (temporal extent/1D) a cover parameter may be used to specify which expected percentage of cover ($\%C_e$) the user requires to obtain a conformant result. By default $\%C_e$ should be set to 100% as full coverage is probably the most useful value.

To illustrate how the percentage of cover parameter is applied, the following example using expected and internal bounding boxes (BB) will be used. Consider A_i the area of the internal BB, A_e the area of the expected BB and A_{int} the area of intersection between both BBs.



Percentage of cover is defined as: $\%C = \left(\frac{A_{int}}{A_e}\right) \times 100$; if $\%C_e$ (defined by the user) is greater than, or equal to, $\%C$ (calculated by the EQE application) then the result is considered conformant/fit (equal to 1). Otherwise it will be considered non-conformant (equal to 0).

The same rationale can be extended to test one-dimensional intervals which apply to temporal extent quality indicator.

Cardinality of comparisons

When multiple values are specified, either in the internal or expected quality indicators, the user must declare how the application shall behave when evaluating multiple match operations. The following options should be available to the user to in order to control multi-value comparisons between internal and expected quality:

- Match all: in this case all specified values must be matched between expected and internal quality;
- Match any: in this case at least one defined value must be matched when comparing expected and internal quality.

For the EQE presented in Section 5.3 of this deliverable D4.5, the option “Match any” was always considered.

Data quality indicators

In this version of the EQE system, we will focus our attention and perform all the quality assessments on a predefined set of quality indicators (QI). The selection of this preliminary set of QI is based on relevant bibliographic references (e.g., Bédard and Vallière, 1995) that provided clear guidelines and criteria for this task. This selection of QI was also based on the assumption that internal quality values for the selected QI could be filled in directly from metadata catalogs (which represented an important limitation in this pilot-study, due to persisting gaps in the metadata information provided for some of the sites). Nevertheless, we argue that the usable set of QI could be extended to include other relevant indicators, in particular those directly related to data and metadata quality description that could be critical to EQE from the user’s perspective (e.g., thematic accuracy and thematic precision, QI that are highly relevant to the evaluation of new datasets to be produced in the BIO_SOS). To implement this functionality, a future extension of the metadata profile should be considered.

The QI selected for this EQE are integrated within the main data characteristics proposed by Bédard and Vallière (1995), as presented in Figure 5.5 and described in Table 5.1.

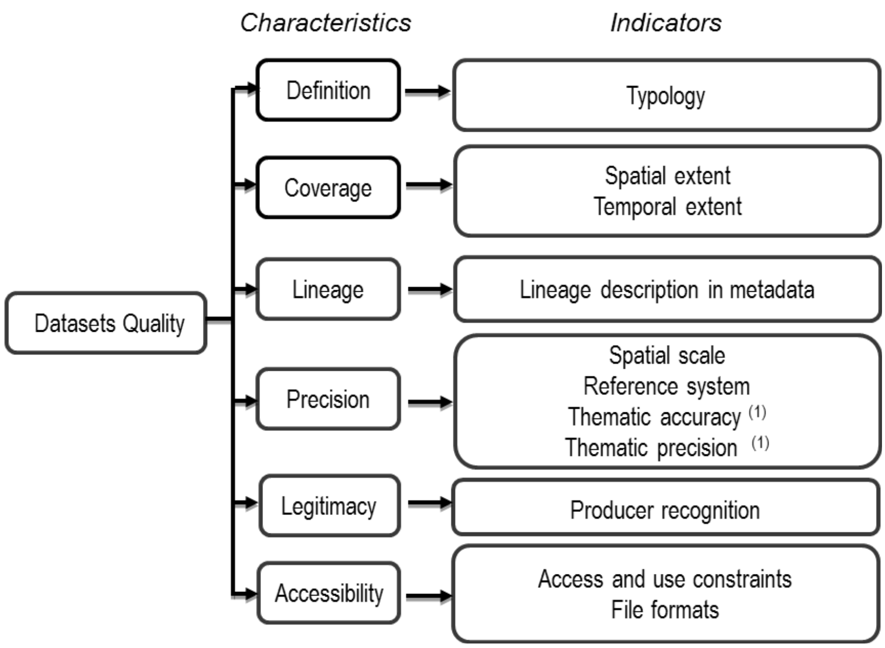


Figure 5.5 – Data characteristics and quality indicators hierarchy. [(1)] Quality indicators of thematic accuracy and thematic precision are intended to use for external quality evaluation of the new products derived throughout the BIO_SOS project.

In the Tables 5.2 - 5.10 we describe in detail the selected QI by defining some characteristics related to:

i) *Name*: provides the designation of the quality indicator;

ii) *Metadata element*: details which metadata element(s) in the current profile will be used to perform comparisons between internal and expected quality;

iii) *Variable type*: defines the data type used in the expected quality indicator and metadata element;

iv) *Allows controlled lists*: determines whether the QI uses a controlled list for metadata fulfillment; the list is either *optional*, i.e. a suggestion of predefined values is suggested to the user but the field admits free text or mandatory, i.e. values can only assume those presented in the controlled list;

v) *Allows multiple values*: specifies whether the QI allows multiple values to be assigned in expected quality; quality indicators derived from metadata elements with cardinality [n] were considered to allow multiple values; if the user chooses to enter multiple values he/she will have to define how the comparisons will be made between internal and expected quality.

vi) *Comparison mechanism*: defines the nature of the comparison between internal and expected quality values.

Table 5.1 – Description of quality indicators in the context of external quality evaluation.

Quality indicator	Element in the metadata editor	Internal/Expected Indicator description
Typology	Classification of spatial data (Classification)	Category of the INSPIRE metadata profile
Spatial extent	Geographic bounding box (Extent)	Bounding box of dataset spatial extent West bound longitude; East bound longitude; South bound latitude; North bound latitude)
Temporal extent	Temporal extent (Extent)	Date or temporal interval of dataset: starting date and ending date
Lineage description in metadata	Lineage (Quality)	Data production methods are described in metadata? [No=0, Yes=1]
Spatial scale	Spatial resolution (Spatial attributes): Equivalent scale or Spatial resolution	Spatial scale of dataset. When the option "spatial resolution" is used, it must consider the same unit of measure for both internal and expected quality matrices
Reference system	Reference system identifier: Code (Spatial reference system)	Alphanumeric value identifying an instance in the namespace, corresponding to the name of reference system (Code EPSG, ESRI or SR.ORG for the reference system)
Thematic accuracy	(1)	Thematic accuracy of dataset (%)
Thematic precision	(1)	Confidence interval of thematic accuracy (%)
Producer recognition	Supplemental Information / Data producer (Identification)	Type of recognition of dataset producer: [Data producer: Official =1; Data producer: Non official=0; Data producer: Unknown =2]
Access and use constraints	Conditions applying to access and use (Constraints)	Type of conditions for access to datasets as described by Article 5(2)(b) and Article 11(2)(f) of Directive 2007/2/EC.
File formats	Distribution format (Distribution)	Type of file formats

(1) Quality indicators of thematic accuracy and thematic precision are intended to use in the external quality evaluation of the new products derived in the BIO_SOS project

Table 5.2 – Description of the quality indicator “Typology”

Attributes	Description (value)
Name	Typology
Metadata element	Topic category
Variable type	Text/string
Allows controlled list	Yes (mandatory)
Admits ranges	No
Allows multiple values	Yes
Comparison mechanism	String comparison

Table 5.3 – Description of the quality indicator “Spatial extent”

Attributes	Description (value)
Name	Spatial extent
Metadata element	Geographic bounding box
Variable type	Geometry box (sequence of 4 coordinates in WGS 1984 GCS)
Allows controlled list	No
Admits ranges	No
Allows multiple values	Yes
Comparison mechanism	Percentage of cover

Table 5.4 – Description of the quality indicator “Temporal extent”

Attributes	Description (value)
Name	Temporal extent
Metadata element	Temporal extent
Variable type	Time interval – start date : end date using YYYY-MM-DD (e.g.: 2011-01-01 : 2011-01-31)
Units	Year Month Day
Allows controlled list	No
Admits ranges	Yes
Allows multiple values	Yes
Comparison mechanism	Percentage of cover

Table 5.5 – Description of the quality indicator “Lineage description”

Attributes	Description (value)
Name	Lineage description
Metadata element	Lineage (free text)

Variable type	Text/character/string
Allows controlled list	No
Admits ranges	No
Allows multiple values	No
Comparison mechanism	Boolean comparison

Table 5.6 – Description of the quality indicator “Reference system”

Attributes	**Description (value)**
Name	Reference system
Metadata element	Reference system identifier (free text)
Variable type	Text/string (suggested format – Entity:SRID, e.g.: EPSG:4326)
Allows controlled list	Yes (optional)
Admits ranges	No
Allows multiple values	Yes
Comparison mechanism	String comparison

Table 5.7 – Description of the quality indicator “Spatial scale”

Attributes	**Description (value)**
Name	Spatial scale
Metadata element	Spatial resolution
Variable type	Integer (equivalent scale) Double (resolution distance) + String (units)
Units	Unitless (equivalent scale) Degrees/ meters (resolution distance; depends on units set on internal quality metadata catalog)
Allows controlled list	No
Admits ranges	Yes
Allows multiple values	Yes
Comparison mechanism	Intersection test

Notes: measurement units in expected quality parameters must match those used in the metadata catalog. To enable the correct fulfillment of this quality indicator, the user must select (from a dropdown list) one of the units available in the metadata residing in the catalog. The user may enter values for one or multiple available units.

Table 5.8 – Description of the quality indicator “Producer recognition”

Attributes	**Description (value)**
Name	Producer recognition

Metadata element	Supplemental information (free text)
Variable type	Text/string
Allows controlled list	Yes (optional)
Admits ranges	No
Allows multiple values	No
Comparison mechanism	String comparison

Table 5.9 – Description of the quality indicator “Access and use restrictions”

Attributes	**Description (value)**
Name	Access and use restrictions
Metadata element	Conditions applying to access and use (free text)
Variable type	Text//string
Allows controlled list	Yes (optional)
Admits ranges	No
Allows multiple values	Yes
Comparison mechanism	String comparison

Table 5.10 – Description of the quality indicator “File formats”

Attributes	**Description (value)**
Name	File formats
Metadata element	Distribution format
Variable type	Text//string
Allows controlled list	Yes
Admits ranges	No
Allows multiple values	Yes
Comparison mechanism	String comparison

5.3 External quality of pre-existing datasets by application context

As already mentioned in sub-section 5.2.2, the methodological framework for EQE can be applied for different data application contexts. The following sub-sections illustrate its application in several application contexts within BIO_SOS, namely: (i) to decide on the usefulness of pre-existing datasets to support habitat mapping within the specified quality requirements for BIO_SOS final habitat maps from EODHaM; (ii) to decide on the usefulness of pre-existing datasets to support sampling design for the collection of new on site data; and (iii) to decide on the usefulness of pre-existing datasets to support the evaluation of the effects of specific pressures on habitats and biodiversity.

Since users' requirements are at the core of the methodological framework, the participation of each Partner was requested for the definition of relevant (and critical) criteria for evaluation and for the definition of expected quality. The expected quality matrix integrating the quality indicators and the core datasets identified for a specific application context (Section 5.2.2) was built according to Table A.2.1 in Appendix IV. A similar matrix (considering the

same datasets, quality indicators and application context) on internal quality was also built based on metadata information (provided in the metadata editor), and according to Table A2.2 in Appendix IV.

5.3.1 Meeting the quality requirements for habitat mapping from EODHaM

Within the collection of pre-existing datasets identified for the several sites in the first stage of Task 4.1 (see **D4.1**), there are datasets potentially useful to support habitat mapping according to the quality requirements in BIO_SOS (as defined by end-users and described in project deliverables **D4.3** and **D5.1**). According to deliverable **D4.4**, available datasets, e.g., archive EO data and products, in situ observational records and maps, and several types of ancillary data (e.g., digital terrain models and cadastral data) can be valuable to support (and also to validate) EO habitat maps.

The core datasets identified by Partners responsible for sites in the UK, Greece, Netherlands and Portugal, for this specific application context, included pre-existing datasets from several topic categories (according to the list of themes in ISO 19115), e.g.: Imagery / Base Maps / Earth Cover (e.g., satellite imagery, air photos); Elevation (e.g., contour maps; digital elevation maps); Transportation (e.g., road networks); Environment (e.g., Natura 2000 network maps); Farming (e.g., LPIS Field information).

In the following case studies we present the results of external quality evaluation for the core datasets identified to support this application context, related to several sites.

Case study nr. 1: PT1 site

Tables 5.11 and 5.12 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT1 site.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 94%, with a standard deviation of 16%. Only one core dataset was considered unfit ("National Road Plan"), due to the quality indicators "Access and use constraints", that showed a mismatch between the values in the internal quality matrix and the expected quality matrix, and "Temporal extent", that presented gaps in metadata information (i.e., in the internal quality matrix). The dataset "ASTER Global Digital Elevation Model (GDEM) [SP]" was classified as partially fit because, although having two quality indicators in non-conformity, these were considered non-critical factors by the user (Table 5.12).

Table 5.11 – External Quality binary matrix for the core (pre-existing) datasets identified to support the production of habitat maps for the PT1 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats

ASTER Global Digital Elevation Model (GDEM) [SP]	1	1	0*	1	0*	1	1	1	1
EO data from Landsat5 TM sensor_19900902 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat4 TM sensor_19910524 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20030805 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20031008 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20070325 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20070731 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20090805 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20100520 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20100707 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat7 ETM+ sensor_20000804 [S]	1	1	1	1	1	1	1	1	1
EO data from Landsat7 ETM+ sensor_20030407 [S]	1	1	1	1	1	1	1	1	1
National Road Plan 2000 [SP]	1	1	0**	0*	0*	0*	1	0**	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Although the average value of fitness for use for the set of core datasets is rather high, it must be highlighted that other datasets are already planned to be acquired within the project (e.g., VHR imagery) in order to meet the quality requirements for habitat mapping in the PT1 site, and that their EQE must be performed before the actual acquisition.

Table 5.12– Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the production of habitat maps for the PT1 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
ASTER Global Digital Elevation Model (GDEM) [SP]	Partially fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the match between values in the internal and expected matrices is partial, but it is within the acceptable range according to the criteria defined by the user; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.

EO data from Landsat5 TM sensor_19900902 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat4 TM sensor_19910524 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20030805 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20031008 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20070325 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20070731 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20090805 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20100520 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20100707 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat7 ETM+ sensor_20000804 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat7 ETM+ sensor_20030407 [S]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
National Road Plan 2000 [SP]	Unfit	44,44	There are several QIs "non-conformant: partially fit": (i) the lineage, because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) the spatial scale, because the value identified in the expected matrix is different from the value in metadata. There are also two QI "non-conformant: unfit" for this dataset: (i) the temporal extent, because the information on this element is not available in the metadata; (ii) the access and use constraints, because none of the values in expected quality matrix match the value in the internal quality matrix.

Case study nr. 2: PT2 site

Tables 5.13 and 5.14 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT2 site.

Table 5.13 – External Quality binary matrix for the core (pre-existing) datasets identified to support the production of habitat maps for the PT2 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.FileFormats
ASTER Global Digital Elevation Model (GDEM) [SP]	1	1	1	1	0*	1	1	1	1

Contour_10m [P]	1	1	1	0*	0*	1	1	0*	1
Elevation points [P]	1	1	1	0*	0*	1	1	0*	1
Geodesic vertex [P]	1	1	1	0*	0*	1	1	0*	1
EO data from Landsat5 TM sensor_20030711 [P]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20070706 [P]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20100425 [P]	1	1	1	1	1	1	1	1	1
EO data from Landsat5 TM sensor_20100730 [P]	1	1	1	1	1	1	1	1	1
EO data from Landsat7 ETM+ sensor_19990708 [P]	1	1	1	1	1	1	1	1	1
EO data from Landsat7 ETM+ sensor_20010526 [P]	1	1	1	1	1	1	1	1	1
EO data from Landsat7 ETM+ sensor_20010915 [P]	1	1	1	1	1	1	1	1	1
Main Road Network [P]	1	1	0**	0*	0*	1	1	0**	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

The EQE for the set of core datasets selected for the present application context reveal a mean fitness for purpose of 87.0%, with a standard deviation of 17.6%. Only one core dataset was considered unfit ("National Road Network"), due to the quality indicators "Access and use constraints", that showed a mismatch between the values in the internal quality matrix and the expected quality matrix, and "temporal extent" that presented gaps in metadata information (i.e., in the internal quality matrix). Once again, the core datasets related to elevation ("ASTER Global Digital Elevation Model (GDEM) [SP]", "Contour_10m [P]", "Elevation points [P]", and "Geodesic vertex [P]") were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.13 and 5.14).

As in the previous case study, although the average values of fitness for use for the set of core datasets is high, it must be highlighted that the acquisition of other datasets is planned (e.g., VHR imagery) to meet the quality requirements for habitat mapping in the PT2 site, and that their EQE must be performed before the actual acquisition (as long as metadata information on these datasets is available).

Table 5.14 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the production of habitat maps for the PT2 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
ASTER Global Digital Elevation Model (GDEM) [SP]	Partially fit	88,89	There is one QIs "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Contour_10m [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
Elevation points [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.

Geodesic vertex [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
EO data from Landsat5 TM sensor_20030711 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20070706 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20100425 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat5 TM sensor_20100730 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat7 ETM+ sensor_19990708 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat7 ETM+ sensor_20010526 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
EO data from Landsat7 ETM+ sensor_20010915 [P]	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
Main Road Network [P]	Unfit	55,56	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata. There is also two QIs "non-conformant: unfit" for this dataset: (i) the temporal extent, because the values in the internal quality matrix and in the expected quality matrix do not match; (ii) the access and use constraints, because none of the value in expected quality matrix does not match the value in the internal quality matrix.

Case study nr. 3: GR sites

Tables 5.15 and 5.16 present the EQE results, including the Boolean matrix (indicating the conformity vs. non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for GR sites. For this case study, the Greek sites were not considered individually, because the available metadata information was not differentiated by site.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 83.3%, with a standard deviation of 7.9%. None of the two identified core datasets was considered unfit and both core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.15 and 5.16).

Table 5.15 – External Quality binary matrix for the core (pre-existing) datasets identified to support the production of habitat maps for GR sites.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
ASTER Global Digital Elevation Model (GDEM) core	1	1	1	1	0*	1	1	1	0*
Natura 2000 core	1	1	0*	1	1	1	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.16 – Summary diagnosis (including diagnosis value and percentage of fit) of the external quality evaluation of the core (pre-existing) datasets identified to support the production of habitat maps for GR sites.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
ASTER Global Digital Elevation Model (GDEM) core	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata; (ii) the file format, because the value identified in the expected matrix is different from the value in metadata.
Natura 2000 core	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the value identified in the expected matrix is different from the value in metadata.

As in the previous case studies, although the average value of fitness for use for the set of core datasets is high, it must be highlighted that other datasets should be acquired (e.g., VHR imagery) to meet the quality requirements for habitat mapping in GR sites, and that their EQE must be performed before the actual acquisition (as long as metadata information on these datasets is available).

Case study nr. 4: UK sites

Tables 5.17 and 5.18 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the UK site. For this case study, the UK sites were not considered individually, because the available metadata information was not differentiated by site.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 67.9%, with a standard deviation of 18.0%. Although several core datasets were considered unfit, this is mainly due to gaps in metadata information (i.e., in the internal quality matrix), and not necessarily due to the low quality of the data. The other datasets were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.17 and 5.18). The low value of fitness for use of the core datasets “OS VM Railways Lines” and “OS VM Roads” (55.6%), partially due to gaps in metadata information, should also be highlighted, since this increases the uncertainty associated to their use.

Table 5.17– External Quality binary matrix for the core (pre-existing) datasets identified to support the production of habitat maps for the UK site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
OS VM Railways Lines	1	1	0*	1	1	0*	0*	1	0*
OS VM Roads	1	1	0**	1	1	0*	0*	1	0*
EA - Welsh Reservoirs	1	1	0**	1	1	0*	0*	1	1
Mean High Water Mark	1	1	0*	1	1	1	0*	1	1
LPIS Field Information	0*	1	0**	1	1	1	0*	1	1
Air photos 2006 Cowi-Vexcel	1	1	1	1	1	1	0*	1	0*

Hyperspectral (EAGLE HAWK) Cors Fochno	1	1	1	1	1	1	0*	1	1
LiDAR data Cors Fochno	1	1	1	1	1	1	0*	1	1
Contours 5m SN	1	1	0*	1	0**	0*	0*	0**	0*

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.18 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the production of habitat maps for the UK site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
OS VM Railways Lines	Partially fit	55,56	There are four QIs "non-conformant: partially fit" for this dataset: (i) the temporal extent, because there is no available information for this element in the metadata; (ii) the spatial scale, because the value in metadata is coarser then the value in the expected matrix; (iii) the producer recognition, because the information on this element is missing in metadata; (iv) the file format, because values in internal and expected quality matrices do not match.
OS VM Roads	Unfit	55,56	There are three QIs "non-conformant: partially fit" for this dataset: (i) the spatial scale, because the value in metadata is coarser then the value in the expected matrix; (ii) the producer recognition, because the information on this element is missing in metadata; (iii) the file format, because values in internal and expected quality matrices do not match. There is one QI "non-conformant: unfit" for this datasets: the temporal extent, because there is no available information for this element in the metadata and this is a CF.
EA - Welsh Reservoirs	Unfit	66,67	There are two QIs "non-conformant: partially fit" for this dataset: (i) the spatial scale, because the value in metadata is coarser then the value in the expected matrix; (ii) the producer recognition, because the information on this element is missing in metadata. There is one QI "non-conformant: unfit" for this datasets: the temporal extent, because there is no available information for this element in the metadata and this is a CF.
Mean High Water Mark	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the temporal extent, because there is no available information for this element in the metadata; (iii) the producer recognition, because the information on this element is missing in metadata.
LPIS Field Information	Unfit	66,67	There are two QIs "non-conformant: partially fit" for this dataset: (i) the typology, because the value in metadata do not match with the value in the expected matrix; (ii) the producer recognition, because the information on this element is missing in metadata. There is one QI "non-conformant: unfit" for this datasets: the temporal extent, because there is no available information for this element in the metadata and this is a CF.

(cont.)

Table 5.18 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the production of habitat maps for the UK site (cont.).

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
Air photos 2006 Cowi-Vexcel	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata; (ii) the file format, because values in internal and expected quality matrices do not match.
Hyperspectral (EAGLE HAWK) Cors Fochno	Unfit	88,89	There is one QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata.
LiDAR data Cors Fochno	Unfit	88,89	There is one QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata.
Contours 5m SN	Unfit	33,33	There are four QIs "non-conformant: partially fit" for this dataset: (i) the temporal extent, because there is no available information for this element in the metadata; (ii) the spatial scale, because the value in metadata is coarser then the value in the expected matrix; (iii) the producer recognition, because the information on this element is missing in metadata; (iv) the file format, because values in internal and expected quality matrices do not match. There are two QI "non-conformant: unfit" for this datasets: (i) the reference system, because there is no available information for this element in the metadata and this is a CF; (ii) the access and use constraints because values in internal quality matrix and expected quality matrix do not match.

Case study nr. 5: NL site

Tables 5.19 and 5.20 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the NL site.

Table 5.19 – External Quality binary matrix for the core (pre-existing) datasets identified to support the production of habitat maps for the NL site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
AHN 2000 - 5m grid (Digital Elevation model)	1	1	1	1	1	1	0*	1	0*
Basic Mapping of Nature 2004	1	1	0**	0*	1	0*	0*	1	0*
Grid 50 x 50 soil map of the Netherlands, scale 1:50000 with peat mapping, version 2006	0**	1	1	1	1	0**	0*	1	0**
Structure mapping of Edese and Ginkelse Heathland 2003	1	1	0**	0*	0**	0**	0*	1	1
Vegetation mapping of Edese and Ginkelse Heathland and Wekeromse Sand 2009	1	1	1	1	0**	0**	0*	1	1
BRT TOP10NL	1	1	1	1	1	0**	0*	0**	0**

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.20 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the production of habitat maps for the NL site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
AHN 2000 - 5m grid (Digital Elevation model)	Partially fit	77.78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata; (iv) the file format, because values in internal and expected quality matrices do not match.
Basic Mapping of Nature 2004	Unfit	44.44	There are four QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the spatial scale, because value in the internal quality matrix is coarser then the value in the expected quality matrix; (iii) the producer recognition, because the information on this element is missing in metadata; (iv) the file format, because values in internal and expected quality matrices do not match. There is one QI "non-conformant: unfit" for this datasets: the temporal extent, because the temporal extent in internal and expected matrices do not match.
Grid 50 x 50 soil map of the Netherlands, scale 1:50000 with peat mapping, version 2006	Unfit	55.56	There is one QI "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata. There area several QIs "non-conformant: unfit" for this dataset: (i) the typology, because values in the internal and expected quality matrices do not match; (ii) the spatial scale, because values in the internal and expected quality matrices do not match (equivalent scale vs resolution distance); (iii) the file format, because values in the internal and expected quality matrices do not match.

Structure mapping of Edese and Ginkelse Heathland 2003	Unfit	44.44	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the producer recognition, because the information on this element is missing in metadata. There are several QIs "non-conformant: unfit" for this datasets: (i) the temporal extent, because the temporal extent in internal and expected matrices do not match; (ii) the reference system, because values in internal and expected quality matrices do not match and is a CF; (iii) the spatial scale, because information is missing in metadata.
Vegetation mapping of Edese and Ginkelse Heathland and Wekeromse Sand 2009	Unfit	44.44	There is one QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata. There are several QIs "non-conformant: unfit" for this datasets: (i) the reference system, because values in internal and expected quality matrices do not match and is a CF; (ii) the spatial scale, because information is missing in metadata.
BRT TOP10NL	Unfit	55.56	There is one QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata. There are several QIs "non-conformant: unfit" for this datasets: (i) the reference system, because because values in internal and expected quality matrices do not match and is a CF; (iii) the spatial scale, because values in the internal and expected qulity matrices do not match (equivalent scale vs resolution distance); (ii) the access and use constraints, because values in internal and expected quality matrices do not match; (iii) the file format, because values in internal and expected quality matrices do not match.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 53.7%, with a standard deviation of 13.0%. Although most of the core dataset were classified as unfit, these results are mainly due to incompatibilities between the filling of the internal quality matrix and of the expected quality matrix, and not necessarily due to the low quality of the data. The dataset “AHN 2000 - 5m grid (Digital Elevation model)” was classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.19 and 5.20).

5.3.2 Supporting sampling designs for new on site campaigns

On site campaigns are intended to provide ground truth data to support calibration and validation of EODHaM habitat maps, but also to provide new information on biodiversity and threats/pressures within the Natura 2000 sites studied in BIO_SOS sites and their surroundings. The collection of ground truth data should be based on a sampling design supported on pre-existing land cover/LCCS/GHC maps, but also on pre-existing information about species distribution and pressures, as described in detail in deliverable **D4.3**.

In the next sub-sections we present and briefly discuss results of the EQE of core datasets intended to support sampling designs for new on site campaigns in several sites.

5.3.2.1 On site campaigns for calibration and validation of EODHaM habitat maps

The calibration and validation of EODHaM habitat maps will be supported by the collection of ground truth data. The sampling design to support the collection of this in situ data will be based on pre-existing datasets, which will increase throughout the project, i.e. new datasets will become available in different stages of the habitat mapping production, allowing adequate support to the calibration and validation tasks (land cover maps available in the beginning of the project → SIAM-derived maps/LCCS maps → GHC maps, as described in deliverable **D4.3**).

In the following case studies we present the results of external quality evaluation for the core datasets identified to support this application context, related to BIO_SOS sites in Portugal, Wales, Netherlands and Greece.

Case study nr. 1: PT1 site

Tables 5.21 and 5.22 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT1 site.

Table 5.21 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the calibration and validation of EODHaM habitat maps for the PT1 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.TemporalExtent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1
COS 1990 [S]	1	1	1	0*	0*	1	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 86.1%, with a standard deviation of 5.6%. None of the core datasets was considered unfit, and all the core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.21 and 5.22).

Although the average value of fitness for use for the set of core datasets is high, it must be highlighted that besides the datasets identified, other datasets should be considered to support the calibration and validation of EODHaM products in PT1 throughout the project, and that their EQE must be timely performed.

Table 5.22 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the calibration and validation of EODHaM habitat maps for the PT1 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [S]	Partialy fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.

Case study nr. 2: PT2 site

Tables 5.23 and 5.24 present the EQE results, including the Boolean matrix (indicating the conformity vs. non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality

matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT2 site.

Table 5.23 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the calibration and validation of EODHaM habitat maps for the PT2 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1
COS 1990 [P]	1	1	1	0*	0*	1	1	1	1
COS 2006 [P]	1	1	1	0*	0*	1	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 84.5%, with a standard deviation of 6.1%. None of the core datasets was considered unfit, and all the core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.23 and 5.24).

As in the previous case study, although the average value of fitness for use for the set of core datasets is high, it must be highlighted that besides the datasets identified, other datasets should be considered to support the calibration and validation of EODHaM products in PT2 throughout the project, and that their EQE must be timely performed.

Table 5.24 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the calibration and validation of EODHaM habitat maps for the PT2 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 2006 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.

Case study nr. 3: GR sites

Tables 5.25 and 5.26 present the EQE results, including the Boolean matrix (indicating the conformity vs. non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the GR sites. For this case study, the Greek sites were not considered individually, because available metadata information was not differentiated by test site.

Table 5.25 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the calibration and validation of EODHaM habitat maps for GR sites.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
ASTER Global Digital Elevation Model (GDEM) core	1	1	1	1	0*	1	1	1	0*
CLC 2000 core	1	1	1	1	1	1	1	1	1
Natura 2000 core	1	1	0*	1	1	1	1	1	1
Orthoimagery GRD 50cm core	1	1	1	0*	1	1	1	0*	0*
EO data (Landsat USGS) core	1	1	1	0*	0*	1	1	1	0*

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

The results of the EQE for the set of core datasets selected for the present application context reveal a mean fitness for purpose of 80.0%, with a standard deviation of 14.5%. None of the core datasets was considered unfit, and the “CLC 2000 core” was classified as fit because all the QIs exhibited conformity between values in the internal quality matrix and in the expected quality matrix. The other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.25 and 5.26).

Table 5.26 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the calibration and validation of EODHaM habitat maps for GR sites.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
ASTER Global Digital Elevation Model (GDEM) core	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata; (ii) the file format, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 core	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
Natura 2000 core	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the match between values in the internal and expected matrices is partial, but it is within the acceptable range according to the criteria defined by the user.
Orthoimagery GRD 50cm core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the access and use constraints, because the values in metadata and in expected quality matrix do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.

EO data (Landsat USGS) core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the reference system in internal and expected quality matrices do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.

As in the previous case studies, although the average value of fitness for use for the set of core datasets is high, it must be highlighted that besides the datasets identified, other datasets might be necessary to support the calibration and validation of EODHaM products in GR sites throughout the project, and that their EQE must be timely performed.

Case study nr. 4: UK sites

Tables 5.27 and 5.28 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the UK site. For this case study, the UK sites were not considered individually, because the available metadata information was not differentiated by site.

Table 5.27 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the calibration and validation of EODHaM habitat maps for the UK site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
1 Km grid	0*	1	0*	1	1	1	0*	1	1
Land Cover 2000	1	1	1	1	1	1	0*	1	0*
Phase 1 Terr. All-Wales.veg	1	1	0**	1	1	1	0*	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.28 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the calibration and validation of EODHaM habitat maps for the UK site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
1 Km grid	Partially fit	55,56	There are four QIs "non-conformant: partially fit" for this dataset: (i) the typology, because the value in metadata do not match with the value in the expected matrix; (ii) the temporal extent, because the temporal interval in metadata do not match with the temporal extent defined in the expected quality matrix; (iii) the producer recognition, because the information on this element is missing in metadata.
Land Cover 2000	Partially fit	88,89	There are two QIs "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata; (ii) the file format, because the value in metadata do not match with the value in the expected matrix.
Phase 1 Terr. AllWales.veg	Unfit	77,78	There is one QI "non-conformant: partially fit" for this dataset: (i) the producer recognition, because the information on this element is missing in metadata. There is one QI "non-conformant: unfit" for this dataset (i) the temporal extent, because the temporal interval in metadata do not match with the temporal extent defined in the expected quality matrix, and this QI in identified as CF

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 74.1%, with a standard deviation of 17.0%. The dataset “Phase 1 Terr. AllWales.veg” was considered

unfit because the temporal extent in metadata information do not include the temporal extent required in the expected quality matrix. The other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.27 and 5.28). Nevertheless, the low value of fitness for use for the core dataset “1Km grid” (55.6%) should be highlighted, since it raises great uncertainty associated with its use.

Once again, as in the previous case studies, although the average value of fitness for use for the set of core datasets is high, it must be highlighted that besides the datasets identified, other datasets might be necessary to support the calibration and validation of EODHaM products in the UK site throughout the project, and that their EQE must be timely performed.

Case study nr. 5: NL site

Tables 5.29 and 5.30 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the NL site.

Table 5.29 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the calibration and validation of EODHaM habitat maps for the NL site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
National Land Use Netherlands - LGN6	0*	1	1	1	1	1	0*	1	0*
Natura 2000 areas as of January 2006	0*	1	1	1	1	0**	0*	1	0*

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.30 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the calibration and validation of EODHaM habitat maps for the NL site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
National Land Use Netherlands - LGN6	Partially fit	66.67	There are two QIs "non-conformant: partially fit" for this dataset: (i) the typology, because the value in metadata do not match with the value in the expected matrix; (ii) the producer recognition, because the information on this element is missing in metadata; (iii) the file format, because values in internal and expected quality matrices do not match.
Natura 2000 areas as of January 2006	Unfit	55.56	There are three QIs "non-conformant: partially fit" for this dataset: (i) the typology, because the value in metadata do not match with the value in the expected matrix; (ii) the producer recognition, because the information on this element is missing in metadata; (iii) the file format, because values in internal and expected quality matrices do not match. There is one QI "non-conformant: unfit" for this datasets: (ii) the spatial scale, because values in the internal and expected quality matrices do not match (equivalent scale vs resolution distance).

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 61.1%, with a standard deviation of 7.9%. Although the dataset “Natura 2000 areas as of January 2006” have been classified as unfit, this result is due to incompatibilities between the filling of the internal quality matrix and of the external quality matrix (in the spatial scale quality indicator), and not necessarily due to the low quality of the data. The dataset “National Land Use Netherlands - LGN6” was classified as partially fit because, although hav-

ing quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.29 and 5.30).

Once again, as in the previous case studies, although the average value of fitness for use for the set of core datasets is high, it must be highlighted that besides the datasets identified, other datasets might be necessary to support the calibration and validation of EODHaM products in the NL site throughout the project, and that their EQE must be timely performed.

5.3.2.2 On site campaigns for collection of new data on biodiversity

Stratifications supporting the sampling design for the collection of new data on biodiversity will be site / pressure / biodiversity indicator dependent. Nevertheless, in most cases the sampling design, as described in deliverable **D4.3**, is based on land cover maps (e.g. LCCS, EODHaM derived products) and on pre-existing species distribution maps.

In the following case studies we present the results of the external quality evaluation for the core datasets identified to support this application context, related to BIO_SOS sites in Portugal and Greece.

Case study nr 1: PT1 site

Tables 5.31 and 5.32 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT1 site.

Table 5.31 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the collection of data on biodiversity for the PT1 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1
COS 1990 [S]	1	1	1	0*	0*	1	1	1	1
ASTER Global Digital Elevation Model (GDEM) [SP]	1	1	0*	1	0*	1	1	1	1
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	1	1	1	1	1	1	1	1	0
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	1	1	0*	1	0*	0**	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.32 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on biodiversity for the PT1 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [S]	Partialy fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
ASTER Global Digital Elevation Model (GDEM) [SP]	Partialy fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the match between values in the internal and expected matrices is partial but it is within the acceptable range, according to the criteria defined by the user; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the file format, because the value identified in the expected matrix is different from the value in metadata.
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	Unfit	66,67	There are two QIs "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the match between values in the internal and expected matrices is partial but it is within the acceptable range, according to the criteria defined by the user; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 82.5%, with a standard deviation of 8.7%. One core dataset was considered unfit (“Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]”) because the spatial scale in metadata is coarser than the spatial scale defined in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.31 and 5.32).

Case study nr. 2: PT2 site

Tables 5.33 and 5.34 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT2 site.

Table 5.33 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the collection of data on biodiversity for the PT2 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1

COS 1990 [P]	1	1	1	0*	0*	1	1	1	1
COS 2006 [P]	1	1	1	0*	0*	1	1	1	1
ASTER Global Digital Elevation Model (GDEM) [SP]	1	1	1	1	0*	1	1	1	1
Contour_10m [P]	1	1	1	0*	0*	1	1	0*	1
Elevation points [P]	1	1	1	0*	0*	1	1	0*	1
Geodesic vertex [P]	1	1	1	0*	0*	1	1	0*	1
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	1	1	1	1	1	1	1	1	0*
Phytosociological associations / vegetation types distribution [SP]	1	1	1	1	0*	0**	1	1	1
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	1	1	1	1	0*	0**	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 79.6%, with a standard deviation of 9.3%. Two core dataset were considered unfit ("Phytosociological associations / vegetation types distribution [SP]" and "Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]") because the spatial scale in metadata is coarser than the spatial scale defined in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.33 and 5.34).

Table 5.34 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on biodiversity for the PT2 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 2006 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
ASTER Global Digital Elevation Model (GDEM) [SP]	Partially fit	88,89	There is one QIs "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Contour_10m [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.

Elevation points [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
Geodesic vertex [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
Vegetation Indices 16-Day L3 Global 250m from Terra MOD-IS sensor (MOD13Q1) [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the file format, because the value identified in the expected matrix is different from the value in metadata.
Phytosociological associations / vegetation types distribution [SP]	Unfit	77,76	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, be-cause the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix
Plant species records / distri-bution data for mainland Por-tugal from phytosociological inventories [SP]	Unfit	77,78	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, be-cause the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix

Case study nr. 3: GR sites

Tables 5.35 and 5.36 present the EQE results, including the Boolean matrix (indicating the conformity vs. non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the GR sites. For this case study, the Greek sites were not considered individually, because available metadata information was not differentiated by test site.

Table 5.35– External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the collection of data on biodiversity for GR sites.

EQM.ResourceTitle	EQM. Ty-pology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lin-eage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 2000 core	1	1	1	1	1	1	1	1	1
Natura 2000 core	1	1	0*	1	1	1	1	1	1
Orthoimagery GRD 50cm core	1	1	1	0*	1	1	1	0*	0*
EO data (Landsat USGS) core	1	1	1	0*	0*	1	1	1	0*

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.36– Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on biodiversity for GR sites.

EQM.ResourceTitle	EQM Dia-gnosis	EQM Percent-age fit	Report
CLC 2000 core	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal qual-ity matrix revealed a total match for all the QI.

Natura 2000 core	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the value identified in the expected matrix is different from the value in metadata.
Orthoimagery GRD 50cm core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the access and use constraints, because the values in metadata and in expected quality matrix do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.
EO data (Landsat USGS) core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the reference system in internal and expected quality matrices do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 80.6%, with a standard deviation of 16.7%. One core dataset was classified as fit ("CLC 2000 core") because all the QIs exhibited conformity between values in the internal quality matrix and in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.35 and 5.36).

5.3.2.3 *On site campaigns for the collection of new data on pressures*

As in the previous application context, stratifications to support the sampling design for the collection of new data on pressures will be site / pressure / biodiversity indicator dependent. Nevertheless, in most cases the sampling design, as described in deliverable **D4.3**, is also based on land cover maps (e.g. LCCS, EODHaM derived products).

In the following case studies we present results of the external quality evaluation for the core datasets identified to support this application context, related to BIO_SOS sites in Portugal and Greece.

Case study nr. 1: PT1 site

Tables 5.37 and 5.38 present the EQE results, including the Boolean matrix (indicating the conformity vs non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT1 site.

Table 5.37 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for the PT1 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1
COS 1990 [S]	1	1	1	0*	0*	1	1	1	1
ASTER Global Digital Elevation Model (GDEM) [SP]	1	1	0*	1	0*	1	1	1	1
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	1	1	1	1	1	1	1	1	0

Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	1	1	0*	1	0*	0**	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.38 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for the PT1 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [S]	Partialy fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
ASTER Global Digital Elevation Model (GDEM) [SP]	Partialy fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the match between values in the internal and expected matrices is partial but it is within the acceptable range, according to the criteria defined by the user; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	Partialy fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the file format, because the value identified in the expected matrix is different from the value in metadata.

(cont.)

Table 5.38 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for the PT1 site *(cont.)*.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	Unfit	66,67	There are two QIs "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the match between values in the internal and expected matrices is partial but it is within the acceptable range, according to the criteria defined by the user; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 82.5%, with a standard deviation of 8.7%. One core dataset was classified as unfit ("Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]") because the spatial scale in metadata is coarser than the spatial scale defined in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.37 and 5.38).

Case study nr. 2: PT2 site

Tables 5.39 and 5.40 present the EQE results, including the Boolean matrix (indicating the conformity vs. non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the PT2 site.

Table 5.39 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for the PT2 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1
COS 1990 [P]	1	1	1	0*	0*	1	1	1	1
COS 2006 [P]	1	1	1	0*	0*	1	1	1	1
ASTER Global Digital Elevation Model (GDEM) [SP]	1	1	1	1	0*	1	1	1	1
Contour_10m [P]	1	1	1	0*	0*	1	1	0*	1
Elevation points [P]	1	1	1	0*	0*	1	1	0*	1
Geodesic vertex [P]	1	1	1	0*	0*	1	1	0*	1
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	1	1	1	1	1	1	1	1	0*
Phytosociological associations / vegetation types distribution [SP]	1	1	1	1	0*	0**	1	1	1
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	1	1	1	1	0*	0**	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.40 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for the PT2 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 2006 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
ASTER Global Digital Elevation Model (GDEM) [SP]	Partially fit	88,89	There is one QIs "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.

Contour_10m [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
Elevation points [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
Geodesic vertex [P]	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata; (iii) access and use constraints, because none of the values in the expected matrix matches with the value metadata.
Vegetation Indices 16-Day L3 Global 250m from Terra MOD-IS sensor (MOD13Q1) [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the file format, because the value identified in the expected matrix is different from the value in metadata.
Phytosociological associations / vegetation types distribution [SP]	Unfit	77,76	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, be-cause the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix
Plant species records / distri-bution data for mainland Por-tugal from phytosociological inventories [SP]	Unfit	77,78	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, be-cause the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 79.6%, with a standard deviation of 9.3%. Two core dataset were considered unfit ("Phytosociological associations / vegetation types distribution [SP]" and "Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]") because the spatial scale in metadata is coarser than the spatial scale defined in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.39 and 5.40).

Case study nr. 3: GR sites

Tables 5.41 and 5.42 present the EQE results, including the Boolean matrix (indicating the conformity vs. non-conformity of each quality indicator, based on the comparison of the internal quality matrix and the expected quality matrix), the criticalness degree for the non-conformant cases, and the diagnosis and report per dataset for the GR sites. For this case study, the Greek sites were not considered individually, because available metadata information was not differentiated by test site.

Table 5.41 – External Quality binary matrix for the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for GR sites.

EQM.ResourceTitle	EQM. Ty-pology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 core	1	1	1	1	1	1	1	1	1
CLC 2000 core	1	1	1	1	1	1	1	1	1
Natura 2000 core	1	1	0*	1	1	1	1	1	1

Orthoimagery GRD 50cm core	1	1	1	0*	1	1	1	0*	0*
EO data (Landsat USGS) core	1	1	1	0*	0*	1	1	1	0*

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.42– Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the sampling design for the collection of data on pressures for GR sites.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 core	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
CLC 2000 core	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
Natura 2000 core	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the value identified in the expected matrix is different from the value in metadata.
Orthoimagery GRD 50cm core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the access and use constraints, because the values in metadata and in expected quality matrix do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.
EO data (Landsat USGS) core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the reference system in internal and expected quality matrices do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 84.5%, with a standard deviation of 16.9%. Two core datasets were classified as fit (“CLC 1990 core” and “CLC 2000 core”) because all the QIs exhibited conformity between values in the internal quality matrix and in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.41 and 5.42).

5.3.3 Assessing the effects of selected pressures on habitats and biodiversity inside and around Natura 2000 sites

This subsection describes case studies of application of EQE to decide on the fitness for use of pre-existing datasets to support the evaluation of the effects of specific pressures on habitats and biodiversity inside and outside the several training and test sites of BIO_SOS.

Case study nr. 1: PT1 *site*

Evaluating the effects of landscape dynamics on key species and habitat types in the PT1 test site

Just downstream of the PT1 test site, a big hydroelectric dam is being built in the Sabor river that will submerge a large extent of the main valley. The new reservoir will eliminate large areas of Annex I habitats and is expected to induce changes in land use patterns in the surrounding slopes and plateau areas. Moreover, a large set of environmental compensatory measures will be implemented and induce further changes in the extent and condition of habitat types as well as on the distribution and abundance of key plant and animal species.

The studies to be developed in the PT1 test site within BIO_SOS will therefore be focused on vegetation dynamics, landscape change and shifts in the occurrence of key species, which are expected to be induced by local socioeconomic change and by environmental compensatory measures. The challenge will be to test whether GHCs maps, derived from EO data (VHR imagery), will (i) accurately distinguish the several habitat types involved in vegetation and landscape dynamics, (ii) adequately represent the spatial structure of landscape mosaics, and (iii) provide a robust connection to the patterns of key plant and animal species. Given the slow rate of change associated to succession processes after abandonment, the general approach will be to survey landscape mosaics with varying composition and spatial structure.

The core datasets identified (Table 5.43) within the collection of pre-existing data for PT1 intend to document land cover and land use as well as local biodiversity, and mainly include pre-existing LC maps and species distribution maps. Further information will be also needed to support the evaluation of the effects of landscape change on key habitats and biodiversity (e.g. GHC/Annex I habitat maps derived within the project and new on site in situ data on pressures and biodiversity).

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 82.2%, with a standard deviation of 9.9%. One core dataset was classified as unfit (”Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]”) because the spatial scale in metadata is coarser than the spatial scale defined in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.43 and 5.44).

Table 5.43 – External Quality binary matrix for the core (pre-existing) datasets identified to support the evaluation of the effects of landscape dynamics on key species and habitat types in the PT1 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
CLC 1990 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2000 [SP]	1	1	1	1	0*	1	1	1	1
CLC 2006 [SP]	1	1	1	1	0*	1	1	1	1
COS 1990 [S]	1	1	1	0*	0*	1	1	1	1
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	1	1	0*	1	0*	0**	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.44 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the evaluation of the effects of landscape dynamics on key species and habitat types in the PT1 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
CLC 1990 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2000 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
CLC 2006 [SP]	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 1990 [S]	Partially fit	77,78	There are two QI "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	Unfit	66,67	There are two QIs "non-conformant: partially fit" for this dataset: (i) the temporal extent, because the percentage of cover is not fully covered but it is within the acceptable range, according to the criteria defined by the user; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix

Case study nr. 2: PT2 site

Modeling the effects of farmland abandonment on species, habitats and landscape heterogeneity in the PT2 test site

Land use change is a major driver of the condition of habitats and biodiversity in marginal farmland of Mediterranean mountains. Among the complex patterns and different pathways currently observed in these landscapes, farmland and pastureland abandonment is expected to drive local landscape homogenization, inducing changes in extent of habitat types as well as the local or regional extinction of ecologically specialized species. This case-study focuses on farmland abandonment in high nature value farmland of the Peneda-Gerês mountain range (PT2 test site). This process has been forecasted to induce profound changes in the landscape levels of vascular plant diversity, including the loss of a considerable number of species in local landscape mosaics and thus a decrease of vascular plant species richness (VPSR) at the landscape level (Lomba 2011).

The results described in Lomba (2011) were based on the statistical analysis of a pre-existing vegetation dataset, so they lack a spatially explicit validation in local landscape mosaics as well as a connection to changes in the extent of habitat types. Therefore, in the context of BIO_SOS, the challenge will be to test whether GHCs maps, derived from EO data (VHR imagery), will (i) accurately distinguish the several habitat types involved in landscape dynamics, (ii) adequately represent the spatial structure of landscape mosaics, and (iii) provide a robust connection to the patterns of key plant species as well as to variations of the VPSR indicator and its additive components (alpha and beta). Given the slow rate of change associated to succession processes after abandonment, the general approach will be to survey landscape mosaics across a gradient of land abandonment.

The core datasets identified (Table 5.45) within the collection of pre-existing data for PT2 intend to document the focal pressure as well as local biodiversity, and mainly include pre-existing LC maps and species distribution maps. Further information will be also needed to support the evaluation of the effects of farmland abandonment on key habitats and biodiversity (e.g. GHC/Annex I habitat maps derived within the project and new on site in situ data on pressures and biodiversity).

Table 5.45 – External Quality binary matrix for the core (pre-existing) datasets identified to support the evaluation of the effects of farmland abandonment on species, habitats and landscape heterogeneity in the PT2 site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
COS 1990 [P]	1	1	1	0*	0*	1	1	1	1
COS 2006 [P]	1	1	1	0*	0*	1	1	1	1
Phytosociological associations / vegetation types distribution [SP]	1	1	1	1	0*	0**	1	1	1
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	1	1	1	1	0*	0**	1	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 77.8%, with a standard deviation of 0%. Two core datasets were classified as unfit ("Phytosociological associations / vegetation types distribution [SP]" and "Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]") because the spatial scale in metadata is coarser than the spatial scale defined in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.45 and 5.46).

Table 5.46 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the evaluation of the effects of farmland abandonment on species, habitats and landscape heterogeneity in the PT2 site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
COS 1990 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
COS 2006 [P]	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage because the information on this element is not available in the metadata; (ii) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Phytosociological associations / vegetation types distribution [SP]	Unfit	77,78	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix
Plant species records / distribution data for mainland Portugal from phytosociological inventories [SP]	Unfit	77,78	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata. There is also one QI "non-conformant: unfit" for this dataset: the spatial scale because the value in the internal quality matrix is coarser than the value in the expected quality matrix

Case study nr. 3: GR sites

Identifying the locations with the highest pressure as well as the locations of greatest conservation interest that fall within the Natura 2000 protected areas

According to deliverable **D2.2**, the main threats and pressures in Greek sites include: (i) in GR1: the water and soil pollution observed in the Kalamas wetland, which is attributed to many human activities, as well as the road construction on the hills that has increased access for hunting and grazing, and the vulnerability of several animal habitats due to a variety of human activities; (ii) in GR2: the restriction of the riparian forest due to the expansion of agricultural land near the river banks, as well as the reduction in grazing contributing to the development of the maquis vegetation and the improvement in its ecological condition, and the endangerment (moderate) of rare species; and (iii) in GR3: the detrimental effects of grazing and arable farming in many animal habitats, along with the hunting and shooting with negative impacts on local fauna, especially migratory birds.

Considering this situation, the main processes in study on GR sites are the extent and change of habitats of European interest, the gains and losses in natural-semi-natural habitats vs. cultivated land, and the fragmentation of natural and semi-natural areas. Partner 2 has selected several core datasets (Table 5.47) within the collection of pre-existing datasets to support the evaluation of the main threats/pressures on natural and semi-natural habitats in GR sites. For this case study, the Greek sites were not considered individually, because available metadata information was not differentiated by test site.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 83.3%, with a standard deviation of 15.3%. Two core datasets were classified as fit ("CLC 1990 core" and "CLC 2000 core") because all the QIs exhibited conformity between values in the internal quality matrix and in the expected quality matrix. All the other core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.47 and 5.48).

Table 5.47 – External Quality binary matrix for the core (pre-existing) datasets identified to support the evaluation of the effects of main pressures on the extent and change of habitats of European interest and on the gains and losses in natural-semi-natural habitats vs. cultivated land and on the fragmentation of natural and semi-natural areas in GR sites.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
ASTER Global Digital Elevation Model (GDEM) core	1	1	1	1	0*	1	1	1	0*
CLC 1990 core	1	1	1	1	1	1	1	1	1
CLC 2000 core	1	1	1	1	1	1	1	1	1
Natura 2000 core	1	1	0*	1	1	1	1	1	1
Orthoimagery GRD 50cm core	1	1	1	0*	1	1	1	0*	0*
EO data (Landsat USGS) core	1	1	1	0*	0*	1	1	1	0*

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.48 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the evaluation of the effects of main pressures on the extent and change of habitats of European interest, on the gains and losses in natural-semi-natural habitats vs. cultivated land and on the fragmentation of natural and semi-natural areas in GR sites.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
ASTER Global Digital Elevation Model (GDEM) core	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata; (ii) the file format, because the value identified in the expected matrix is different from the value in metadata.
CLC 1990 core	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
CLC 2000 core	Fit	100,00	The comparison between the values in the expected quality matrix and in the internal quality matrix revealed a total match for all the QI.
Natura 2000 core	Partially fit	88,89	There is one QI "non-conformant: partially fit" for this dataset: (i) the reference system, because the value identified in the expected matrix is different from the value in metadata.
Orthoimagery GRD 50cm core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the access and use constraints, because the values in metadata and in expected quality matrix do not match; (iv) the file format, because the value identified in the expected matrix is different from the value in metadata.
EO data (Landsat USGS) core	Partially fit	66,67	There are several QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because this information is missing in metadata; (ii) the reference system in internal and expected quality matrices do not match; (iii) the file format, because the value identified in the expected matrix is different from the value in metadata.

Eventually, further datasets (acquired or produced) will also be needed to support the evaluation of the effects in study, and in that case, the EQE of those data should be timely performed.

Case study nr. 4: UK sites

Evaluating the impact of changing hydrological regimes on the dynamics of the active raised bog and estuarine environment and the impacts of bog restoration and invasive species

The main pressures and/or threats in the UK site concerns the active raised bog, which is particularly vulnerable to changes in the hydrological regime caused by intentional drainage but also by flooding and saline intrusion associated with storm surges and sea level rise, as detailed in deliverable **D2.2**. Additionally, succession and encroachment (e.g., by woodland, by *Molinia caerulea*) can potentially lead to a reduction in the area and function of the raised bog, and control is largely through grazing and active management. Invasive species, including *Rhododendron*, also occur, and may lead to a depletion of native plant species. Much of the bog is surrounded by agricultural land and airborne (e.g., nitrogen) deposition or direct fertilisation may favour certain plant species (e.g., *Molinia*) at the detriment of others, as also detailed in deliverable **D2.2**.

Considering this situation, the main processes in study on the UK site concern the evaluation of the impact of changing hydrological regimes on the dynamics of the active raised bog and estuarine environment and the impacts of bog restoration and invasive species. Partner 11 selected three core datasets (Table 5.49) from the collection of pre-existing datasets to support the evaluation of those impacts on raised bog. For this case study, the UK sites were not considered individually, because the available metadata information was not differentiated by site.

Table 5.49 – External Quality binary matrix for the core (pre-existing) datasets identified to support the evaluation of the impact of changing hydrological regimes on the dynamics of the active raised bog and estuarine environment and the impacts of bog restoration and invasive species in the UK site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats
Protected Sites: National Nature Reserves - Final Bounds - GIS Dataset	1	1	1	0*	1	1	0*	1	1
Sites (SAC)	1	1	1	0*	1	1	0*	1	0*
SSSI Qualifying features	1	1	1	0*	1	1	0*	1	1

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Table 5.50 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support the evaluation of the impact of changing hydrological regimes on the dynamics of the active raised bog and estuarine environment and the impacts of bog restoration and invasive species in the UK site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
Protected Sites: National Nature Reserves - Final Bounds - GIS Dataset	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (iii) the producer recognition, because the information on this element is missing in metadata.
Sites (SAC)	Partially fit	66,67	There are three QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (iii) the producer recognition, because the information on this element is missing in metadata; (iii) the file format, because values in internal and expected quality matrices do not match.
SSSI Qualifying features	Partially fit	77,78	There are two QIs "non-conformant: partially fit" for this dataset: (i) the lineage, because there is no available information for this element in the metadata; (iii) the producer recognition, because the information on this element is missing in metadata.

The results of the EQE for the set of core datasets selected for this application context reveal a mean fitness for purpose of 74.1%, with a standard deviation of 6.4%. All the core datasets selected were classified as partially fit because, although having quality indicators in non-conformity, these were considered non-critical factors by the user (Tables 5.49 and 5.50).

Case study nr. 5: NL site

Evaluating the impacts on airborne nitrogen deposition and other pressures on habitat biodiversity

The Dutch study area (the Veluwe), which is the largest end moraine in Netherlands, falls under the Habitats Directive as well as under the Birds Directive. Although the inland sand dunes of the Veluwe are still among the largest in Europe, their area has severely diminished, as described in detail in deliverable **D2.2**. According to this deliverable, the most important pressure to the biodiversity on the main habitats in this area is nitrogen deposition caused by intensive agriculture in the region. Partner 4 selected one core dataset (Table 5.51) from the collection of pre-existing datasets to support the evaluation of the impacts the main pressures on habitats biodiversity.

Table 5.51 – External Quality binary matrix for the core (pre-existing) datasets identified to support the evaluation of the impact of pressures in habitat biodiversity in the NL site.

EQM.ResourceTitle	EQM. Typology	EQM.Spatial Extent	EQM.Temporal Extent	EQM. Lineage	EQM.Ref System	EQM.Sp Scale	EQM.Prod Recognition	EQM.Access UseRestrictions	EQM.File Formats

National Vegetation Releves Database - LVD	0*	1	1	1	1	0**	0*	1	0

Values of 1 correspond to conformities, while values of 0 correspond to non-conformities. For the non-conformity cases, the degree of criticalness is represented as: * – partially conformant, i.e., the values in the internal and expected quality matrices do not match but the quality indicator is not identified as critical factor; ** – non-conformant, i.e., the values in the internal and expected quality matrices do not match and the quality indicator is identified as critical factor.

Although the dataset "National Vegetation Releves Database - LVD" has been classified as unfit, this result is due to incompatibilities between the filling in in the internal quality matrix and in the external quality matrix (in the spatial scale quality indicator), and not necessarily due to the low quality of the data (Tables 5.51 and 5.52).

Table 5.52 – Summary diagnosis (including diagnosis value and percentage of fit) of EQE of the core (pre-existing) datasets identified to support evaluation of the impact of pressures in habitat biodiversity in the NL site.

EQM.ResourceTitle	EQM Diagnosis	EQM Percentage fit	Report
National Vegetation Releves Database - LVD	Unfit	55.56	There are two QIs "non-conformant: partially fit" for this dataset: (i) the typology, because values in internal and expected quality matrices do not match; (ii) the producer recognition, because the information on this element is missing in metadata; (iii) the file format, because values in internal and expected quality matrices do not match. There is one QI "non-conformant: unfit" for this datasets: (i) the spatial scale, because values in the internal quality matrix are coarser then the value in the expected quality matrix.

Eventually, further datasets (acquired or produced) will also be needed to support the evaluation of the effects in study, and in that case, the EQE of those data should be timely performed.

5.4 Final remarks

As previously mentioned, the determination of uncertainty derived from the use of pre-existing datasets, inferred by the external quality evaluation methodology proposed/used for (core) pre-existing datasets, intends to help avoiding using data potentially inappropriate to solve users' problems or needs. The application of the EQE framework for core datasets selected per application context by the several Partners highlighted the importance of providing accurate and complete information on metadata. The use of a complete metadata profile, as proposed in the second stage of Task 4.1 (see Section 5.2.1), allowed a detailed characterization of spatial data in use within the BIO_SOS project, thus supporting well-informed decisions on data usability in different application contexts.

Although the methodology has only been applied here to core datasets, we emphasize that it can be used for all datasets identified in the first stage of Task 4.1 (as detailed in deliverable **D4.1**) as well as for new datasets to be produced or acquired within the project (e.g. GHC/Annex I habitat maps, new field data, etc.). This is particularly useful for quality control of BIO_SOS data products, thus enforcing a continuous and iterative cycle of quality evaluation, improvement and validation. For this purpose, the use of a complete metadata profile is paramount to allow a complete description of spatial data elements. Complementarily, we strengthen that an accurate fulfillment of metadata, in particular for new datasets to be acquired or produced throughout the project, is crucial, and that the best time to collect metadata is while data are being produced (when risks of providing less accurate information and the costs of searching for information are lower), as highlighted elsewhere (e.g. Federal Geographic Data Committee 2000).

Concerning the products to be generated by the EODHaM system, we argue that it is particularly important to include and deliver to end-users detailed metadata regarding accuracy (as highlighted in deliverable **D6.3**) as part of the external quality evaluation of datasets. Therefore, the inclusion of elements related with thematic accuracy and precision in the BIO_SOS metadata profile (following the specifications of ISO 19115 for data quality elements) is

already in study as a follow-up of Task 4.1. The integration of these elements in the metadata editor will provide the required information for their use as quality indicators, further enhancing the informative and decision-support power of this EQE framework.

For the purpose of this deliverable, the EQE was implemented in a pilot or alpha version (in R language), although the EQE methodology has already been structured to directly integrate the WebGIS platform currently under development. This is possible due to the modular architecture of the platform (see Section 4.a), which facilitates the future integration of this (and other) external application(s). Therefore, the EQE methodology will soon be fully available in the WebGIS platform, allowing Partners to use it, either to perform evaluation of data quality under specified application contexts, or to perform a ranked search of available datasets based on EQE concepts.

6. GEOPORTAL DEVELOPMENT AND IMPLEMENTATION PLAN

6.1 Implementation plan

The development of the BIO_SOS WebGIS platform was influenced by two specific processes/phases that constrained its progress, namelly: (i) collecting metadata from all the Partners in charge of training and test sites; and (ii) the definition of the methodologies as well as the design and the implementation of the technologies that were used. The future exploration and optimization of the results and the continuity and sustainability of the platform should involve governance policies, strategies and practices at the different levels and components (data, technologies, users, rules and policies).

The development process of the BIO_SOS WebGIS platform and metadata geoportal within the time frame of Task 4.1 has included four phases:

i) requirements and platform specification, involving consultation and participation of all members and project Partners (development);

ii) programming and integration of the different components of the WebGIS platform, implementation and operation;

iii) installation and platform loading with real data (for test and evaluation); and

iv) platform functioning assessment by team members of all Partners participating in the task (e.g. through on-line reports).

The implementation plan beyond the delivery of D4.5 (end of month 12), running at least until month 20 of the project, intends to optimize the BIO_SOS WebGIS platform and metadata geoportal by:

i) incorporating new (metad)data (produced within the project) and the results of ongoing evaluation; and

ii) implementing iteratively the four development phases listed above, according to project and platform development cycles.

In the context of a platform expansion, maintenance and sustainability plan, to be submitted to the consortium, new (meta)data will be uploaded and there will be advances in the analysis and editing modules, as well as in the implementation of functional integration with other WebGIS platforms and applications. In the months following the end of Task 4.1, project Partners should expand the use of the WebGIS platform and, in parallel with other tasks, perform tests on the several platform functionalities, including: (i) implementation of new query features, spatial data processing and data sharing with other platforms; (ii) loading and management of new data and metadata; and (iii) testing and assessment tools to monitor the content and functioning of the platform.

The current and expected platform operation development will require operational adjustment, proposal, discussion and finally adoption of a BIO_SOS WebGIS platform sustainability plan, including:

i) *documentation and project communication*, highlighting the importance of comprehensive and detailed documentation describing each phase and element, but also of developing mechanisms for promoting dissemination to internal and external Partners that may facilitate the adoption of project products;

ii) *platform improvement*, in order to share, locate and access the several relevant documents (e.g. technical documents, training materials, test suites, implementation references and guidelines);

iii) *identification of users' training needs and organization of training processes* that promote individual and institutional capacity building centered in simple WebGIS platform handling and management, as well as training activities aimed at developing competences on collecting, identifying, organizing and producing spatial metadata, conducting spatial data quality evaluations and metadata management associated to spatial modeling processes;

iv) *discussion and internalization of a data and metadata collection responsibility chain*, as well as rules for metadata fulfillment, quality assessment, coordination, administration and platform management;

v) *increasing the usefulness of the results obtained in the project*, through integration and data sharing with other information systems, platforms and/or projects that result from similar initiatives that promote knowledge networks;

vi) *data quality and metadata evaluation and management*, which requires the application of standards for internal quality evaluation, but also monitoring the quality of data (pre-existing and to be produced) for different purposes compared to the existing normative and legal framework; and

vii) *intellectual property registration and management* regarding project resources and products used in the implementation and operation of the platform, namely collected and produced (meta)data, considering the public funding of the project.

The future sustainability of the platform will depend upon its use and usefulness identified by decision-makers, researchers, technicians and generalist users, and will be strongly influenced by system usability issues, particularly the easiness of management and data sharing between the different platforms or implementation stages of the BIO_SOS project. The design, development and testing of site interoperability, data access and maintenance procedures will be important at this last stage of development of the information system. Nominated data custodians, for example, should already be maintaining and updating data and publishing metadata associated with the data archives for which they have accepted responsibility. The future usefulness, continuous adjustment and system optimization are dependent of the WebGIS platform test and monitoring (see Section 6.2), public access and, network integration dynamics, namely the evolution of interoperability and data sharing.

6.2 Operation and continuous evaluation

The development of the information system includes the implementation and functioning of the WebGIS platform, which will gain performance with its continued use and with further network connections and services, but also with the implementation and harmonization of quality evaluation procedures, client configurations and future capabilities. The main future concerns with the WebGIS platform are related to the usability, performance and quality of network services, as the collected datasets need to migrate, as easy, fast and securely as possible, from server to client. Therefore, in order to maximize functionality and performance of the WebGIS platform, it is necessary to apply, test and monitor actions based on adequate methods, tools and real or simulated processes in order to define development and optimization techniques.

The approval and launch of the BIO_SOS WebGIS platform and metadata geoportal should correspond to an extended period of formal tests, which implies real-context operation tests, namely of performance, usability and network connection, as well as service quality evaluation. This phase implies a first conformance test (SDI-cookbook, 2008), in which each project product is to be tested in order to determine its conformity in relation to its implementation objectives.

The tests proposed for the WebGIS platform assume the whole information system already developed (users, technology, organization, and functions) and are related to: (i) content and data quality, and (ii) function capability and

quality of the associated user-centered services, as well as both client and server-side test analysis perspectives. Server-side testing will use web server log files and exploratory data analysis (Markov et al., 2007). Recommended metrics include number of visits, actions, session duration, relationship between visit actions and session, average time per page, duration for individual pages.

The test and monitoring activities will be centered on performance, usability and quality network services, including different methods and process, namely: (i) developing scenarios (consisting of groups of tasks that should be performed by the user within a specified time); (ii) implementing and analysing use-case(s) (specific and representative activities and requests oriented towards system content and functions and to users' needs); (iii) developing on-line general and specific questionnaires related to users' actions and decisions; (iv) developing comparative tests using similar web information system platforms; (v) simulating multiple virtual users accessing the platform concurrently, based on real users' spatial, temporal and thematic request patterns; (vi) using web site test tools which permit and facilitate the testing of performance, load and stress; and (vii) considering European and other international legal Web platform testing frameworks.

Performance, reliability or volume testing

The performance, reliability or volume testing integrates: (i) load and performance testing, namely modeling the expected use of a software program by simulating multiple users accessing the program concurrently; the load generator imitates the behavior of actual real users; and (ii) stress test functionality, in which load is raised beyond normal use patterns, in order to test system response at unusually high or peak loads (Abugessaisa and Östman, 2011).

The performance qualities of the WebGIS platform are measured in terms of system ability to respond to the users' operations and queries, and, in addition, in terms of the reliability of the service over time related to three components, namely the client, the server and the network. Response times are used as measurable performance indicators, calculated as the total time (as perceived by the user) between sending a particular request and obtaining a response or the result of the query or any other service (e.g. download services). Reliability is based on measurable indicators, which require statistical data that can be obtained from a user or a system log book and are related with availability. The system availability is taken as a function of the mean time between failure and the mean time to restore the services (Abugessaisa and Östman, 2011).

Performance tests will therefore include the application of metrics related with: (i) response time (average request processing time, maximum request processing time, minimum request processing time, and overall performance); and (ii) errors (percentage of errors and average bandwidth). The protocol to be adopted defines variables like test duration, number of users and other particular questions (e.g. the time specified to simulate the time taken by a real user to take action before clicking to the next action, connection speed, broadband).

Usability

The usability evaluation of geographic information services applies the ISO Standards framework, namely: (i) ISO 9241-11: Guidance on Usability; (ii) ISO 9241-210: Human-centered design for interactive system; and (iii) ISO 20282-3: Test method for measuring usability. Usability is the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-11: 1998). At the same time: (i) effectiveness is defined as the accuracy and completeness with which users achieve specified goals (ISO 9241-11:1998); (ii) efficiency refers to the resources expended in relation to the accuracy and completeness with which users achieve goals (ISO 9241-11:1998); (iii) and satisfaction is defined as the freedom from discomfort, and positive attitudes towards the use of the product (ISO 9241-11:1998, definition 3.4).

These usability and acceptance tests assist in measuring the extent to which the system complies with the usability design procedures (Nielsen, 2000) and also contribute to maintain a usable and acceptable system (Figure 6.1). The model assumes usability according to three dimensions: (i) content of the system; (ii) user satisfaction; and (iii) ease of interaction. The evaluation of users' satisfaction should consider user interface design, content testing, and functional capabilities by different individual or collective project users.

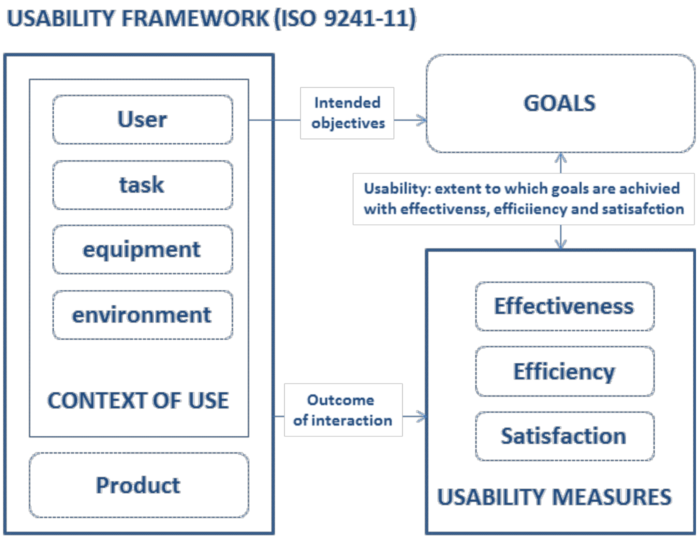


Fig 6.1 - Usability framework (ISO 9241-11).

Usability measures will include: (i) effectiveness (level of completion and errors); (ii) efficiency (time and cognitive workload); and (iii) satisfaction (through psychometric questionnaire). The indicators/statistics produced to evaluate the BIO_SOS WebGIS platform usability will imply understanding the goal(s) for which the tool was designed and the context(s) of use of the several tests, and can integrate information about: (i) layout, visual clarity; (ii) consistency; (iii) navigating; (iv) terminology; (v) feedback, (vi) user control and help; (vii) functionality; (viii) error handling; (ix) design weakness; and (x) general opinion. This usability evaluation integrates periodic questionnaire based on user profiles, uses requisites, platform response, as well as suggestions and practice proposals supported on users' personal opinions.

In addition, and considering the analyses of user´s interaction with the metadata platform, the system will create a repository with information on operations made by users, registering data about the type of operation executed (e.g. search, edit,…) and other relevant information. This data repository will be a valuable source of information that will allow to understand user´s needs and to identify the interest of platform functionalities for Partners, and therefore to support future developments of the metadata geoportal.

Network connection and service quality evaluation

The potential network connection and services available for the BIOSOS WebGIS platform will require monitoring in order to expand and optimize data, metadata and service sharing between internal applications and other WebGIS applications and platforms.

The Regulation of Network Services (Annex I; INSPIRE 2008) defines usual evaluation criteria, which will be adopted in the case of the BIO_SOS WebGIS platform: (i) time delay, which is the time between the moment when something is initiated, and the moment when its effects begin or become detectable; (ii) error occurrence, e.g. aver-

age time between consecutive failures; percentage of errors occurred during a session or some iteration of requests; (iii) availability, i.e. the percentage of customers that can access a Web-based; and (iv) quality of service, regarding performance (in periods out of peak load), capacity , and availability.

Along the project we expect to develop and integrate in the WebGIS platform a tool enabling the online publishing of simple metrics and statistics, as well as internal and external reports and graphs regarding platform tests and activities. This platform testing and monitoring is intended to manage users´ profiles and access in order to learn about individual and institutional use patterns of the several geoportal functions under an optimization and sustainability framework.

7. SYNTHESIS AND FINAL REMARKS

The WebGIS platform of BIO_SOS assumes a collaborative dimension, in its development and functioning, and a modular and evolving dimension as a guarantee of its utility and continuity. This WebGIS platform intends to improve the internal and external communication in the context of the project, and to promote the quality evaluation of relevant datasets for specific modeling tasks as well as of the datasets resulting from image processing and classification. The analysis of requirements and their specification allowed: (i) the definition and characterization of different user access profiles; (ii) the proposal and implementation of a metadata profile and standardized metadata editor; and (iii) the generation, management and operation of a geographic viewer within a metadata geoportal.

The external quality evaluation of core pre-existing datasets was based on a comparison between the intrinsic characteristics of the data (i.e. their internal quality) and the characteristics of the data as required by users (i.e. their expected quality). The several case studies reporting results of the evaluation of such core pre-existing datasets per application context and site/country revealed a medium to high average value of fitness for use for the studied set of core datasets. Nonetheless, besides the core pre-existing datasets identified by Partners, other datasets (both pre-existing and newly acquired or produced) may be necessary to support different application contexts throughout the project, and their quality evaluation must be timely performed. The results also emphasize the importance of an accurate filling of metadata information, which should also be a concern for new datasets to be acquired or produced throughout the project. Therefore, the continuous and iterative character of quality evaluation is once again stressed, towards the improvement and validation of datasets for specific purposes.

Future BIO_SOS governance should consider continuous internal and external quality evaluation of all pre-existing and newly produced spatial datasets, assuming a single common quality framework for the project (see project management and quality assessment plan - Task 8 and deliverable **D8.5**). Spatial data quality evaluation and management throughout the project will imply: (i) transparent user profiles and access management in order to monitor individual and institutional patterns of use of the several WebGIS functions, besides contributing to the integrity and security of the system and consequently to data quality management; (ii) user capacity-building promotion through training courses and acquisition of practical experience; and (iii) development of tools and procedures for spatial data quality evaluation, management, monitoring and internal reporting. Future platform development requires the adoption of a BIO_SOS WebGIS sustainability plan that includes: (i) documentation and project communication; (ii) platform improvement related to sharing, locating and accessing relevant datasets; (iii) training of users; (iv) implementation of a data and metadata collection responsibility chain; (v) communication and sharing of new project results and products; (vi) data quality and metadata evaluation and management; and (vii) intellectual property registration and management.

Developed between months 3 and 12 of BIO_SOS, Task 4.1 has contributed to a set of key issues for the future workflow in the project, since it has provided: (i) a description and evaluation of relevant and potentially useful pre-existing datasets; (ii) informed support to the identification of important data gaps and to the establishment of priorities for new dataset acquisition; (iii) a set of guidelines and tools for data quality evaluation and management throughout the project; and (iv) a collaborative platform and a metadata geoportal for the consortium. The data quality evaluation procedures established in Task 4.1, and implemented as quality evaluation tools in the geoportal, are of high importance for future work in BIO_SOS. Beyond the specific context and objectives of BIO_SOS, this quality evaluation of pre-existing datasets across several European and non-European countries provided a formal

assessment of the actual usefulness of a wide range of habitat, biodiversity and ancillary datasets to support or frame the monitoring of habitats, biodiversity and landscapes under international goals, targets and indicators.

8. LIST OF ACRONYMS

BB	Bounding box
BIO-SOS	Biodiversity Multisource Monitoring System: from Space to Species
BR	Brazil site
CF	Critical Factors
CQL	Contextual Query Language
CSW	Catalogue Service for the Web
DBMS	Database Management System
DS	Discovery service
EO	Earth Observation
EODHaM	Earth Observation Data for Habitat Monitoring
EPSG	European Petroleum Survey Group
EQE	External Quality Evaluation
FTP	File Transfer Protocol
GHC	General Habitat Categories
GML	Geography Markup Language
GR	Greece sites
GUI	Graphical User Interface
ISO	International Organization for Standardization
KML	Keyhole Markup Language
LC	Land cover
LCCS	Land cover classification system
MS	Milestone
NL	Netherlands site
OGC	Open Geospatial Consortium
PT1	Portuguese Sabor-Maçãs site
PT2	Portuguese Peneda-Gerês site
QI	Quality indicators
SDI	Spatial Data Infrastructure
SIAM	Satellite Image Automatic Mapper™
SLD	Styled Layer Descriptor
SQL	Structured Query Language
SRID	Spatial Reference system Identifier
UK	United Kingdom sites
VHR	Very high resolution
WebGIS	World Wide Web Geographic Information System
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Packages
XML	Extensible Markup Language

9. Appendices

APPENDIX I

Guidelines for selecting and describing core datasets

Selection of core datasets

The BIO_SOS WebGIS collaborative platform, currently under implementation in Task 4.1 (due by November 2011), will integrate a metadata geoportal that will allow all Partners to systematize and publish metadata on pre-existing datasets as well as on newly collected or produced data products concerning the different sites in the project.

The inventory of pre-existing spatial datasets implemented in the first stage of Task 4.1 (Deliverable 4.1) was based on the characterization of databases for each test site through a simplified metadata profile. The information then collected provided a brief description of the available datasets from each site, allowing geoportal users to consult and discover information on topic category, spatial resolution, temporal extent, date of publication geographic bounding box, file type, author, property, and spatial reference system.

In this second stage of Task 4.1, more detailed information – complete metadata profile – shall also be collected and subsequently included in the metadata geoportal, but only for "core datasets" (also called "fundamental datasets"). Fields in this complete metadata profile will include information on dataset identification, classification, related keywords, geographic location, temporal reference, quality and validity, constraints related to access and use, responsible party, reference system, and data quality information, thus allowing the enhancement of data discovery functionalities.

Starbuck (2001) defines "core dataset" as the 'core' or 'minimum' set of spatial information needed to support national and subregional environmental assessment, decision making, and environmental reporting. Nerbert (2004) also defines "core data" as a set of geographic information that is necessary for optimal use of most GIS applications. Gyamfi-Aidoo et al. (2006) define "fundamental geospatial datasets" as the "minimum primary set of data that cannot be derived from other datasets, and that are required to spatially represent phenomena, objects, or themes (…) at local, national, sub-regional and/or regional scales".

Among the many criteria and guidelines for identifying fundamental geospatial datasets (adapted from Gyamfi-Aidoo et al. 2006; RDM Working Group 2002), these should:

I. include a complete coverage over the area of interest;
II. include sufficient detail (considering the requirements of different application contexts);
III. have acceptable standards and validation processes that ensure consistency, reliability, quality, continuity and accuracy (since these datasets are those upon which other thematic datasets and products (e.g. from in situ data) will be developed);
IV. provide a context to allow others to better understand the information that is being presented;

V. be collected by official agencies;

VI. enable the merging of data from various sources i.e. integration of data from heterogeneous sources in a network environment, which requires a neutral format and an underlying conceptual model; and

VII. allow that a diversity of users from different sectors may derive significant benefit from their use.

Considering these definitions, and as previously described in Deliverable D4.1, the identification of core/fundamental datasets within the collection of pre-existing datasets from each BIO_SOS site should support decisions on a wide range of application contexts within the scope of the project (i.e. habitat and biodiversity monitoring). Therefore, for the BIO_SOS purpose, the operational concept of core/fundamental datasets can be described as the most relevant datasets for (i) the assessment of the process(es) in study by each Partner at each site, and (ii) the overarching goals of BIO_SOS, and thus their selection should identify the fundamental datasets considering several application contexts:

- production of habitat maps within the specified quality requirements for BIO_SOS (as defined by end-users and described in D4.3 and D5.1);
- sampling design for the collection of new on site campaigns for: (i) calibration or validation datasets of EODHaM habitat maps; (ii) collection of new data on biodiversity; (iii) collection of new data on pressures; and
- evaluation of the effects of specific pressures on habitats and biodiversity.

Considering the diversity of sites and their characteristics, as well as the diversity of processes/pressures being studied, the fundamental geospatial datasets need to be identified within appropriate user-defined frameworks, and thus contribution from site partners is necessary for the identification of core datasets for each site.

Metadata are a key element of those core/fundamental datasets, making data understandable and sharable among users over time. Therefore, the filling in of a complete metadata profile for the core datasets identified by BIO_SOS Partners is necessary for the study of pressures, while allowing detailed search and support sharing mechanisms in the geoportal.

<u>Selection of core datasets – an example</u>

An example of the procedure / steps for selecting the core datasets within the application context related to the evaluation of the effects of specific pressures on habitats and biodiversity in the PT2 test site (evaluation of the effects of meadow abandonment on key habitats and biodiversity indicators) is presented in Figure 1.

For these core datasets, it will be necessary to fulfill a complete metadata profile in the geoportal (according to the information in the next section), and later an evaluation of their external quality will be performed according to the methodology presented in deliverable D4.1.

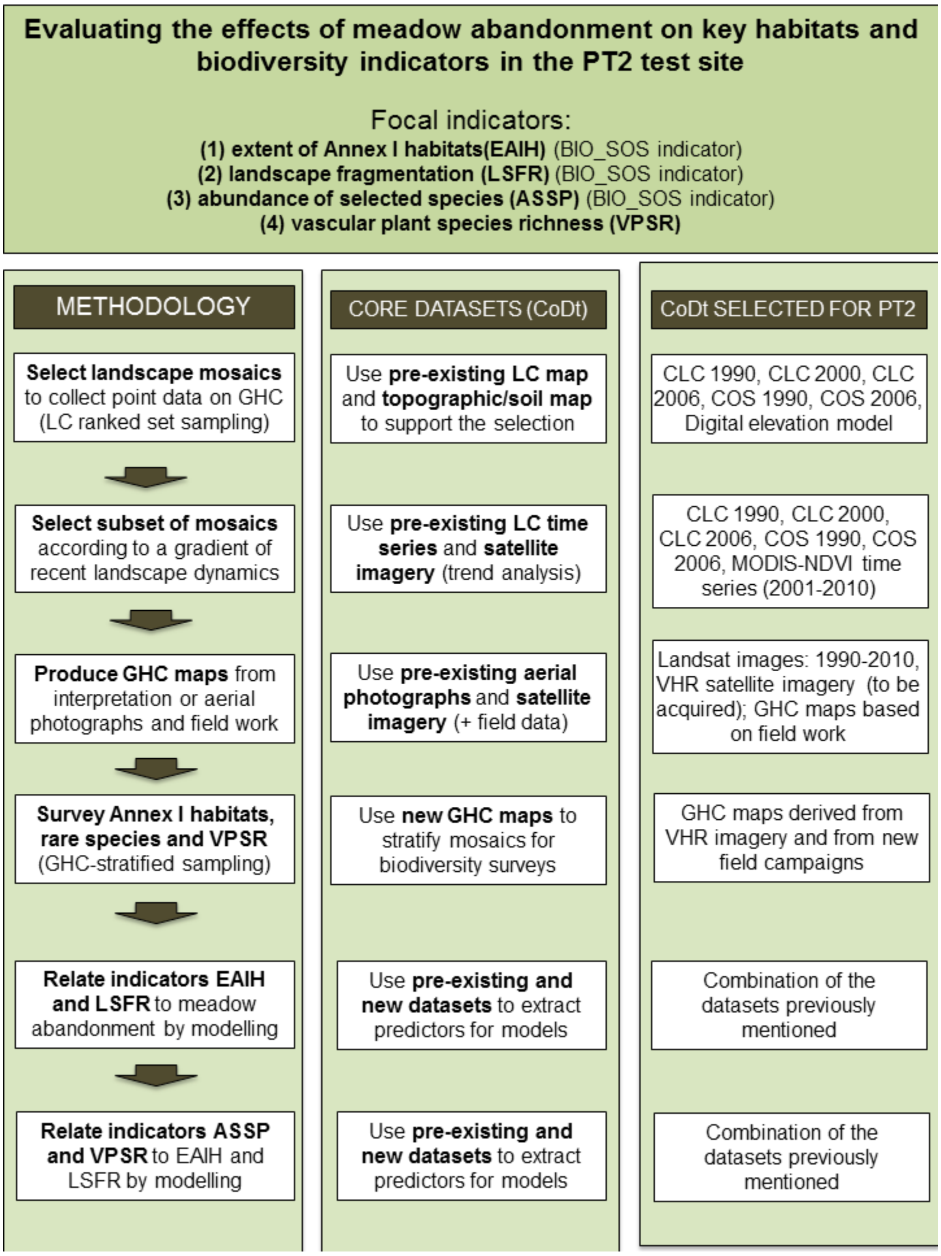


Figure 1 – Core datasets to select for the evaluation of the effects of meadows abandonment on key habitats and biodiversity indicator in PTs test site

Complete metadata profile on core (pre-existing) datasets

The complete metadata profile on core datasets (selected within the collection of pre-existing datasets identified for each test site by partners) is compliant with INSPIRE profile, which take account of relevant international standards, namely the standards EN ISO 19115 and EN ISO 19119. Therefore, the complete metadata profile follow the concepts, themes and framework of the INSPIRE metadata regulation with a few additional inputs (essentially for data quality assessment), which are considered in the standard ISO 19115.

The complete metadata table will be completed by each test-site partner for the selected core datasets, following a harmonized completion procedure that includes:

I. the confirmation of the inexistence of absent information;

II. the confirmation of the inexistence of duplicates;

III. the confirmation of the inexistence of invalid characters or categories;

IV. the validation of the coordinate systems names and acronyms; and

V. the validation of the classification of each dataset according to the INSPIRE Directive.

In the following sections, a definition of concepts is provided to support the filling in of the complete metadata profile, based on European Commission Joint Research Centre (2009) and ISO (2003).

The metadata fields marked with an asterisk (*) are mandatory.

1. Metadata on metadata

*1.1 Metadata language **

This is the language in which the metadata elements are expressed (Figure 1). The value domain of this metadata element is limited to the official languages of the Community expressed in conformity with ISO 639-2. The value domain of this metadata element is a Code List.

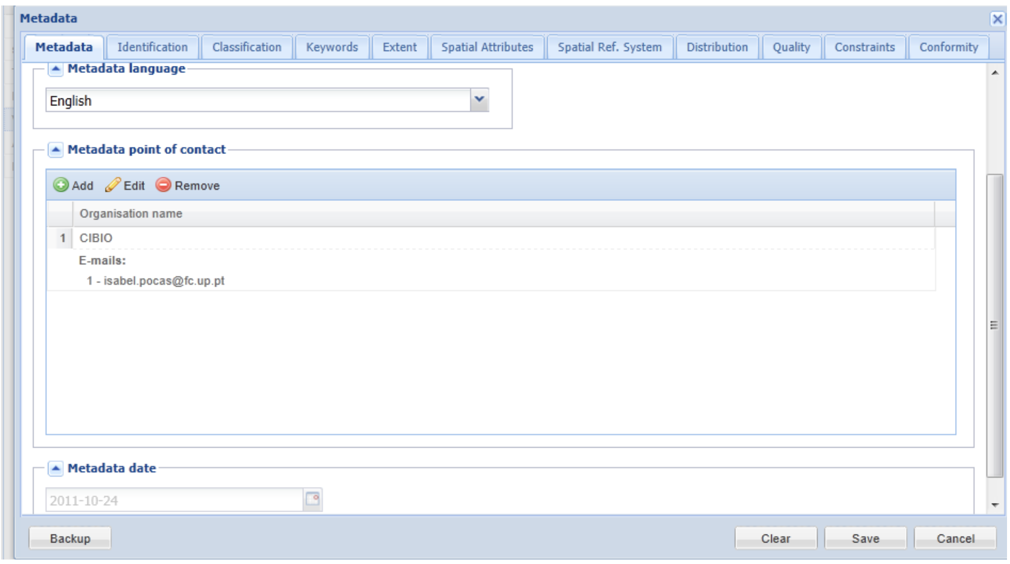


Figure 1 – Metadata on metadata on the editor

1.2 Metadata point of contact

This is the description of the organisation responsible for the creation and maintenance of the metadata. This description shall include:

*1.2.1 Organisation name**

Name of the organisation responsible for the creation and maintenance of the metadata (Figure 1).

*1.2.2 Email**

A contact e-mail adress of the organisation responsible for the creation and maintanance of metadata (e-mail address as a character string)

1.3 Metadata date

The date which specifies when the metadata record was created or updated (Figure 1). This date shall be expressed in conformity with ISO 8601 (e.g.: 2011-09-14).

2. Dataset Identification

*2.1 Resource title **

Characteristic, and often unique, name by which the resource is known (Figure 2). The value domain of this metadata element is free text.

2.2 Date of publication

This is the date of publication of the resource when available, or the date of entry into force.

2.3 Date of last version

This is the date of last revision of the resource, if the resource has been revised. There shall not be more than one date of last revision.

*2.4 Date of creation **

This is the date of creation of the resource. There shall not be more than one date of creation.

2.5 Identifier - Unique Resource Identifier

*2.5.1 Code **

Value uniquely identifying the resource (Figure 2). The value domain of this metadata element is a mandatory character string code, generally assigned by the data owner (e.g., code: http://image2000.jrc.it#image2000_1_nl2_multi (URI form); code: 527c4cac-070c-4bca-9aaf-92be-ce7be902 (UUID form)).

2.5.2 Namespace

Character string namespace uniquely identifying the context of the identifier code (for example, the data owner).

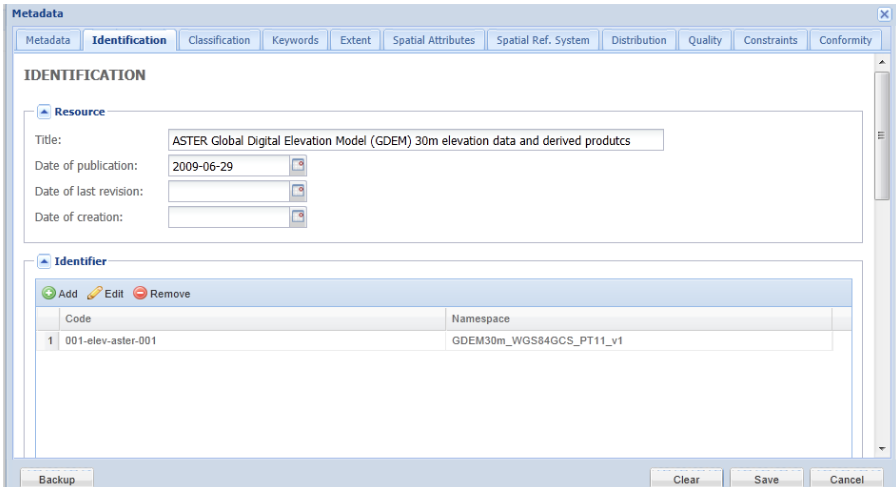


Figure 2 – Dataset identification on the metadata editor

2.6 Resource abstract *

Brief narrative summary of the content of the resource (Figure 3). The value domain of this metadata element is free text.

2.7 Responsible party

This is the description of the organisation responsible for the establishment, management, maintenance and distribution of the resource (Figure 3). This description shall include the name of the organisation as free text and a contact e-mail address as a character string, as indicated in the following fields.

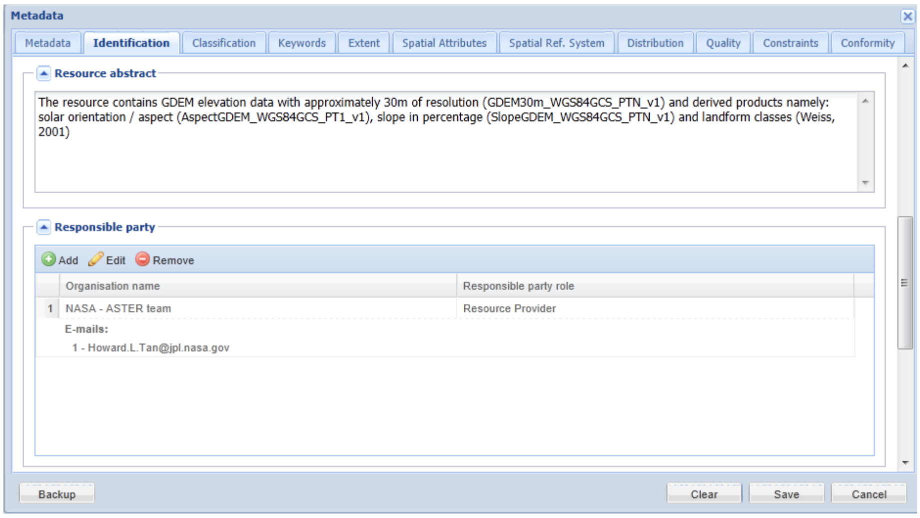


Figure 3 – Dataset identification (resource abstract and dataset responsible party) on the metadata editor

2.7.1 Organization name *

Name of the organisation responsible the establishment, management, maintenance and distribution of the resource. The domain of this metadata element is free text

*2.7.2 Organisation email **

A contact e-mail address of the organisation responsible for the establishment, management, maintenance and distribution of the resource (e-mail address as a character string).

*2.7.3 Responsible party role**

This is the role of the responsible organisation. The value domain of this metadata element is a Code List (ISO 19115 – Annex B5.5):

- Author - party who authored the resource;
- Custodian - party that accepts accountability and responsibility for the data and ensures appropriate care and maintenance of the resource;
- Distributor - party who distributes the resource;
- Originator - party who created the resource;
- Owner - party that owns the resource;
- Point Of Contact - party who can be contacted for acquiring knowledge about or acquisition of the resource;
- Principal Investigator - key party responsible for gathering information and conducting research;
- Processor - party who has processed the data in a manner such that the resource has been modified;
- Publisher - party who published the resource;
- Resource Provider - party that supplies the resource;
- User - party who uses the resource;

*2.8 Resource language **

The language(s) used within the resource (Figure 4). The value domain of this metadata element is limited to the languages defined in ISO 639-2. The value domain of this metadata element is a Code List.

2.9 Supplemental Information

Any other descriptive information about the dataset. Information of the recognition of the dataset producer shall be included (official *vs* non-official) (Figure 4). The domain of this metadata element is free text.

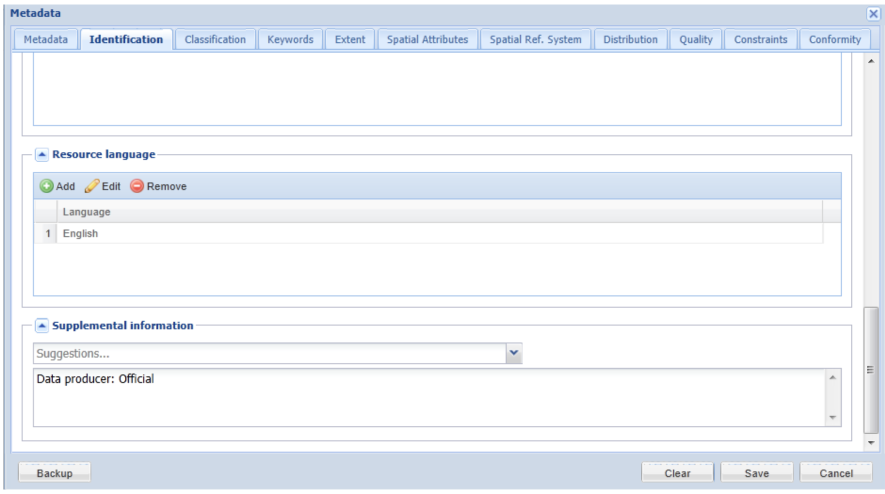


Figure 4 – Dataset identification (resource language and supplemental information on the dataset) on the metadata editor

3. Classification of spatial data

*3.1 Topic category **

Main theme(s) of the dataset (Figure 5). The topic category is a high-level classification scheme to assist in the grouping and topic-based search of available spatial data resources. Main theme(s) of the dataset. The topic category is a high-level classification scheme to assist in the grouping and topic-based search of available spatial data resources.

The list of themes is (ISO 19115):

- *Biota* - flora and/or fauna in natural environment (e.g., wildlife, vegetation, biological sciences, ecology)
- *Boundaries* - legal land descriptions (e.g., political and administrative boundaries);
- *Climatology/Meteorology/Atmosphere* - processes and phenomena of the atmosphere (e.g., cloud cover, weather, climate);
- *Economy* - economic activities, conditions and employment (e.g., production, labour, revenue, commerce, industry);
- *Elevation* - height above or below sea level (e.g., altitude, bathymetry, digital elevation models, slope)
- *Environment* - environmental resources, protection and conservation (e.g., environmental pollution, waste storage and treatment);
- *Farming* - rearing of animals and/or cultivation of plants (e.g., agriculture, irrigation, aquaculture, plantations, herding);
- *Geoscientific Information* - information pertaining to earth sciences (e.g., geophysical features and processes, geology, erosion);
- *Health* - health, health services, human ecology, and safety (e.g., disease and illness, factors affecting health, hygiene);

- *Imagery/BaseMaps/Earth Cover* - Base maps (e.g., land cover, topographic maps, imagery, unclassified images)
- *Inland Waters* - inland water features, drainage systems and their characteristics (e.g., rivers and glaciers, salt lakes, dams, water quality);
- *Intelligence/Military* - military bases, structures, activities (e.g., barracks, training grounds, military transportation);
- *Location* - positional information and services (e.g., addresses, geodetic networks, control points);
- *Oceans* - features and characteristics of salt water bodies (excluding inland waters) (e.g., tides, tidal waves, coastal information);
- *Planning / Cadastre* - information used for appropriate actions for future use of the land (e.g., land use maps, zoning maps, cadastral surveys, land ownership);
- *Society* - characteristics of society and cultures (e.g., settlements, anthropology, archaeology, education, demographic data, census information);
- *Structure* - man-made construction (e.g., buildings, museums, churches, factories, housing, monuments);
- *Transportation* - means and aids for conveying persons and/or goods (e.g., roads, airports/airstrips, railways);
- *Utilities / Communication* - energy, water and waste systems and communications infrastructure and services (e.g., hydroelectricity, geothermal, solar and nuclear sources of energy)

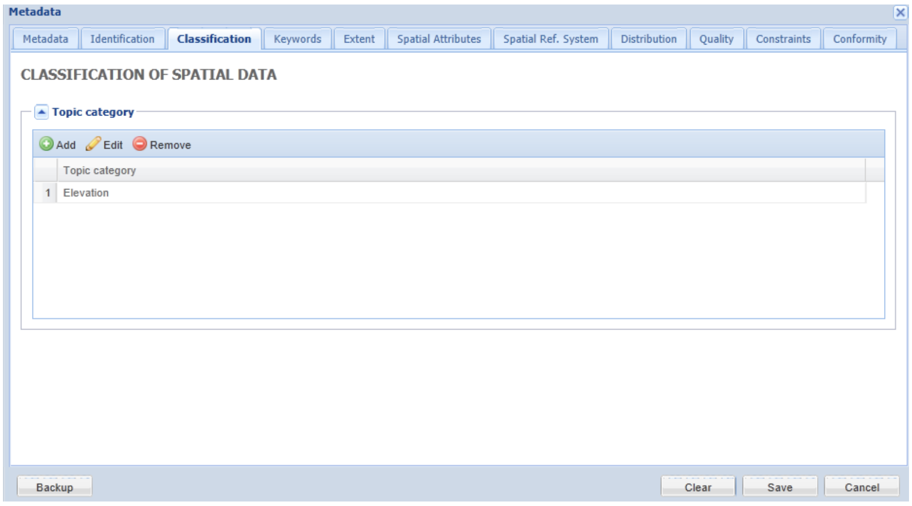


Figure 5 – Topic category of the dataset

4. Keyword

Commonly used word, formalised word or phrase used to describe the subject. While the topic category is too coarse for detailed queries, keywords help narrowing a full text search and they allow for structured keyword search.

*4.1 Keyword from INSPIRE data themes **

Commonly used word, formalised word or phrase used to describe the subject. The value domain of this metadata element is an INSPIRE thematic category from the controlled list (Figure 6a). The list of themes (as described in the Annex I, II and II of the Directive INSPIRE) originating from the general environmental multilingual thesaurus (GEMET; http://www.eionet.europa.eu/gemet/inspire_themes) is:

- *Addresses* - Location of properties based on address identifiers, usually by road name, house number, postal code;
- *Administrative units* - Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries;
- *Agricultural and aquaculture facilities* - Farming equipment and production facilities (including irrigation systems, greenhouses and stables;
- *Area management / restriction / regulation zones & reporting units* - Areas managed, regulated or used for reporting at international, European, national, regional and local levels. Includes dumping sites, restricted areas around drinking water sources, nitrate-vulnerable zones, regulated fairways at sea or large inland waters, areas for the dumping of waste, noise restriction zones, prospecting and mining permit areas, river basin districts, relevant reporting units and coastal zone management areas;
- *Atmospheric conditions* - Physical conditions in the atmosphere. Includes spatial data based on measurements, on models or on a combination thereof and includes measurement locations;
- *Bio-geographical regions* - Areas of relatively homogeneous ecological conditions with common characteristics;
- *Buildings* - Geographical location of buildings;
- *Cadastral parcels* - Areas defined by cadastral registers or equivalent;
- *Coordinate reference systems* - Systems for uniquely referencing spatial information in space as a set of coordinates (x, y, z) and/or latitude and longitude and height, based on a geodetic horizontal and vertical datum;
- *Elevation* - Digital elevation models for land, ice and ocean surface. Includes terrestrial elevation, bathymetry and shoreline;
- *Energy resources* - Energy resources including hydrocarbons, hydropower, bio-energy, solar, wind, etc., where relevant including depth/height information on the extent of the resource;
- Environmental monitoring facilities - Location and operation of environmental monitoring facilities includes observation and measurement of emissions, of the state of environmental media and of other ecosystem parameters (biodiversity, ecological conditions of vegetation, etc.) by or on behalf of public authorities;
- *Geographical grid systems* - Harmonised multi-resolution grid with a common point of origin and standardised location and size of grid cells;
- *Geographical names* - Names of areas, regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest;

- *Geology* - Geology characterised according to composition and structure. Includes bedrock, aquifers and geomorphology;
- *Habitats and biotopes* - Geographical areas characterised by specific ecological conditions, processes, structure, and (life support) functions that physically support the organisms that live there. Includes terrestrial and aquatic areas distinguished by geographical, abiotic and biotic features, whether entirely natural or semi-natural;
- *Human health and safety* - Geographical distribution of dominance of pathologies (allergies, cancers, respiratory diseases, etc.), information indicating the effect on health (biomarkers, decline of fertility, epidemics) or well-being of humans (fatigue, stress, etc.) linked directly (air pollution, chemicals, depletion of the ozone layer, noise, etc.) or indirectly (food, genetically modified organisms, etc.) to the quality of the environment;
- *Hydrography* - Hydrographic elements, including marine areas and all other water bodies and items related to them, including river basins and sub-basins. Where appropriate, according to the definitions set out in Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (2) and in the form of networks;
- *Land cover* - Physical and biological cover of the earth's surface including artificial surfaces, agricultural areas, forests, (semi-)natural areas, wetlands, water bodies;
- *Land use* - Territory characterised according to its current and future planned functional dimension or socio-economic purpose (e.g. residential, industrial, commercial, agricultural, forestry, recreational);
- *Meteorological geographical features* - Weather conditions and their measurements; precipitation, temperature, evapotranspiration, wind speed and direction;
- *Mineral resources* - Mineral resources including metal ores, industrial minerals, etc., where relevant including depth/height information on the extent of the resource;
- *Natural risk zones* - Vulnerable areas characterised according to natural hazards (all atmospheric, hydrologic, seismic, volcanic and wildfire phenomena that, because of their location, severity, and frequency, have the potential to seriously affect society), e.g. floods, landslides and subsidence, avalanches, forest fires, earthquakes, volcanic eruptions;
- *Oceanographic geographical features* - Physical conditions of oceans (currents, salinity, wave heights, etc.);
- *Orthoimagery* - Geo-referenced image data of the Earth's surface, from either satellite or airborne sensors;
- *Population distribution* – demography - Geographical distribution of people, including population characteristics and activity levels, aggregated by grid, region, administrative unit or other analytical unit.;
- *Production and industrial facilities* - Industrial production sites, including installations covered by Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control (1) and water abstraction facilities, mining, storage sites;

- *Protected sites* - Area designated or managed within a framework of international, Community and Member States' legislation to achieve specific conservation objectives;
- *Sea regions* - Physical conditions of seas and saline water bodies divided into regions and sub-regions with common characteristics;
- *Soil* - Soils and subsoil characterised according to depth, texture, structure and content of particles and organic material, stoniness, erosion, where appropriate mean slope and anticipated water storage capacity;
- *Species distribution* - Geographical distribution of occurrence of animal and plant species aggregated by grid, region, administrative unit or other analytical unit.;
- *Statistical units* - Units for dissemination or use of statistical information;
- *Transport networks* - Road, rail, air and water transport networks and related infrastructure. Includes links between different networks. Also includes the trans-European transport network as defined in Decision No 1692/96/EC of the European Parliament and of the Council of 23 July 1996 on Community Guidelines for the development of the trans-European transport network (1) and future revisions of that Decision;
- *Utility and Government services* - iincludes utility facilities such as sewage, waste management, energy supply and water supply, administrative and social governmental services such as public administrations, civil protection sites, schools and hospitals.

4.2 Keyword from online repositories

Commonly used word, formalised word or phrase used to describe the subject.

The value domain of this metadata element is a keyword(s) of a controlled list within the options: GEMET Concepts; GEMET Groups; GEMET Themes; GEOSS Societal Benefit Areas; INSPIRE Feature Concept Dictionary; INSPIRE Glossary; ISO 19119 geographic services taxonomy (see the keyword tab in URL: http://www.inspire-geoportal.eu/index.cfm/pageid/342 for each one of the presented options) (Figure 6b).

4.3 Free keyword

*4.3.1 keyword value**

Commonly used word, formalised word or phrase used to describe the subject.

The value domain of this metadata element is free text.

4.3.2 Originating controlled vocabulary

Name of the formally registered thesaurus or a similar authoritative source of keywords. This fields is only mandatory If the keyword value originates from a controlled vocabulary (thesaurus, ontology), e.g. GEMET. The citation of the originating controlled vocabulary shall be provided in the following fields:

*4.3.2.1 Title **

Title associated to the originating controlled vocabulary list or publication (e.g., GEMET - INSPIRE Spatial Data Themes)

*4.3.2.2 Reference date **

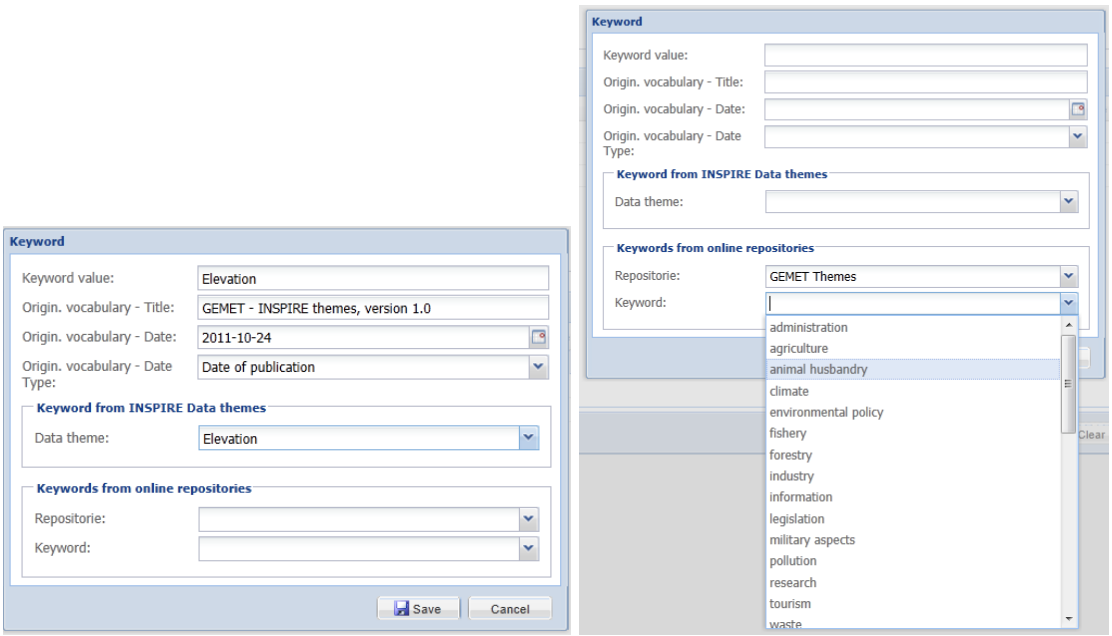
Reference date associated to the originating controlled vocabulary list or publication (e.g., for GEMET - INSPIRE Themes version 1.0: 2008-06-01)

*4.3.2.3 Date type **

Date type associated to the originating controlled vocabulary list or publication (e.g., for GEMET - IN-SPIRE Spatial Data Themes: publication).

The value domain of this metadata element is a Code List:

- Date of creation;
- Date of last revision;
- Date of publication.



(a) *(b)*

Figure 6 – Keyword on metadata editor: (a) Keyword from INSPIRE data themes; (b) Keyword from online repositories

5. Extent

*5.1 Geographic bounding box **

This is the extent of the resource in the geographic space, given as a bounding box. The bounding box shall be expressed with westbound and eastbound longitudes, and southbound and northbound latitudes in decimal degrees, with a precision of at least two decimals. (e.g.: -7.81 -4.97 40.83 42.73) (Figure 7).

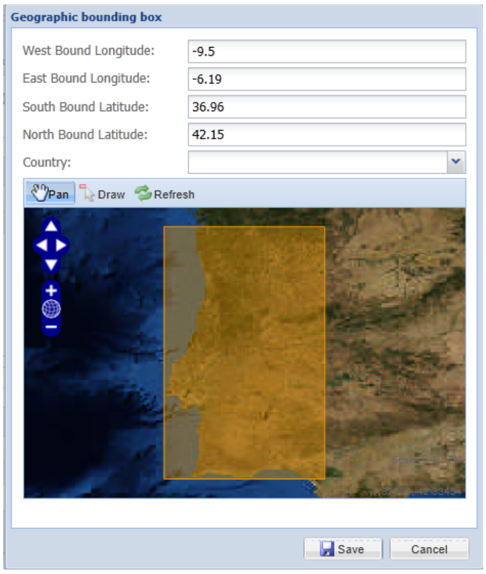


Figure 7 – Geographic bounding box of the dataset on the metadata editor

5.2 Temporal extent

Time period covered by the content of the resource. This time period may be expressed as an interval of dates expressed through the starting date and end date of the interval (in cases of individual date, starting date and ending date are the same) (Figure 8).

*5.2.1 Starting date**

Starting date of the time period (e.g., 2000-01-01)

*5.2.2 Ending date**

Ending date of the time period (e.g., 2010-12-31)

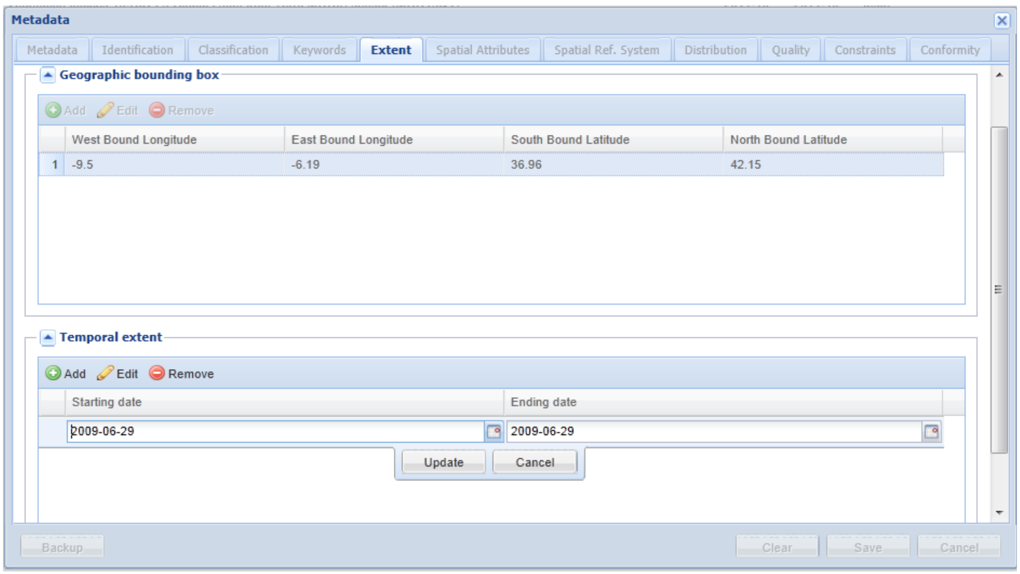


Figure 8 – Temporal extent of the dataset on the metadata editor

6. *Spatial attributes*

6.1 Spatial representation type

Method used to spatially represent geographic information (Figure 9). The domain of this metadata element is a Code List (ISO 19115 – Annex B5.26):

- Vector – vector data is used to represent geographic data;
- Grid - grid data is used to represent geographic data;
- Text/Table - textual or tabular data is used to represent geographic data;
- Tin - triangulated irregular network;
- Stereo Model - three-dimensional view formed by the intersecting homologous rays of an overlapping pair of images;
- Video - scene from a video recording.

6.2 Spatial resolution

Level of detail of the data set expressed as a set of zero to many resolution distances (typically for gridded data and imagery-derived products) or equivalent scales (typically for maps or map-derived products) (Figure 9). The value domain of this metadata element is free text.

6.2.1 Equivalent scale

An equivalent scale is generally expressed as an integer value expressing the scale denominator.

6.2.2 Resolution distance

A resolution distance shall be expressed as a numerical value associated with a unit of length.

6.2.3 Unit of measure

Unit of measure of the distance value. Mandatory if resolution distance field is used to describe to resource.

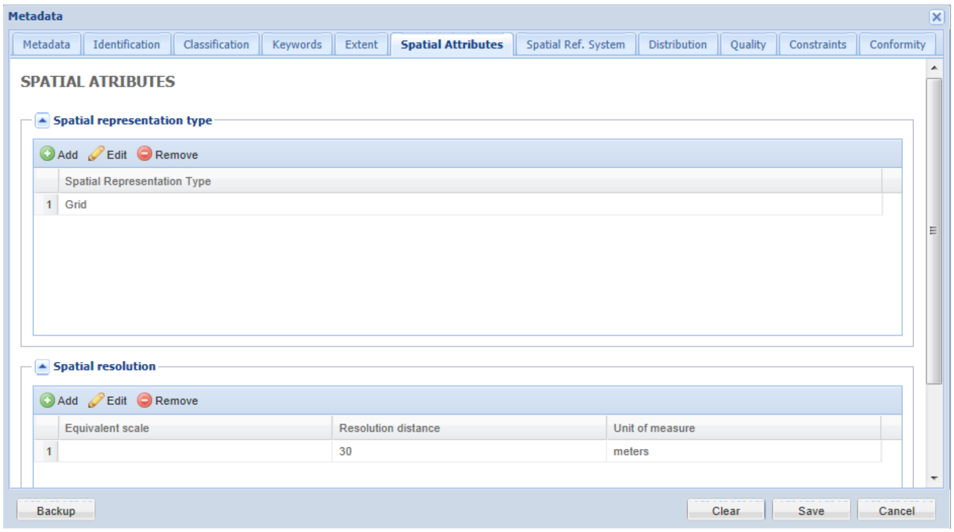


Figure 9 – Spatial attributes, including spatial representation type and spatial resolution, on the metadata editor

7. Spatial reference system information

Description of the spatial reference systems used in the dataset.

*7.1 Reference System Identifier **

Name of reference system. Information on reference system (code and codespace) can be obtained in http://spatialreference.org/

*7.1.1 Code **

Alphanumeric value identifying an instance in the namespace (Figure 10). The value domain of this metadata element is free text. (e.g. EPSG:23029 (ED50/UTM zone 29N); EPSG: 32629 (WGS 84 / UTM zone 29N)).

7.1.2 CodeSpace

Name or identifier of the person or organization responsible for namespace. The value domain of this metadata element is free text but is advisable to use the Code List:

- EPSG
- ESRI
- SR-ORG

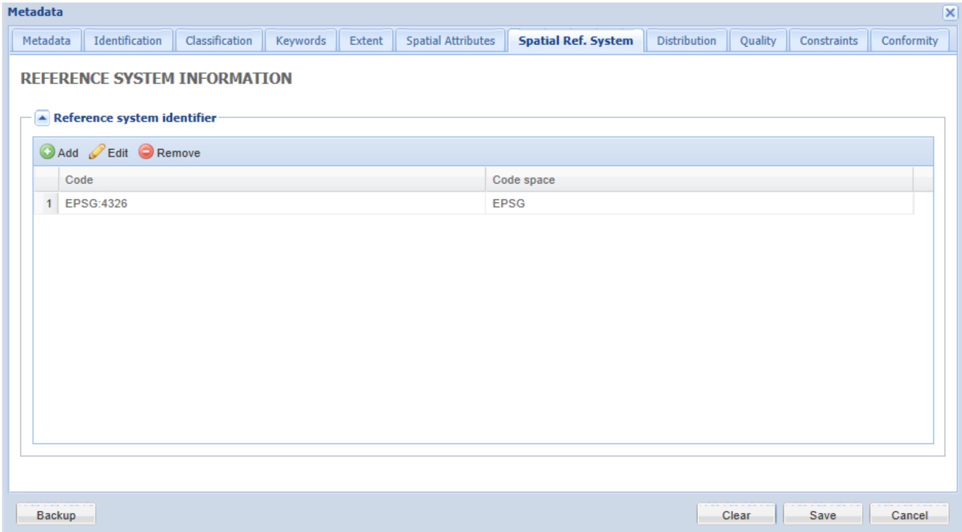


Figure 10 – Reference system information on the metadata editor

8. Distribution Information

8.1 Format Name *

Name of the data transfer format(s). The value domain of this metadata element is free text but is advisable to use the Code List proposed in the metadata editor (e.g. Raster formats: e.g., GeoTIFF (.tif); HFA - Erdas Imagine (.img); MrSID - Multi-resolution Seamless Image Database; ENVI - ENVI .hdr Labelled Raster; …; Vector formats: e.g., ESRI Shapefile - ESRI Shapefile; GML - GML; PostgreSQL - PostgreSQL; …) (Figure 11).

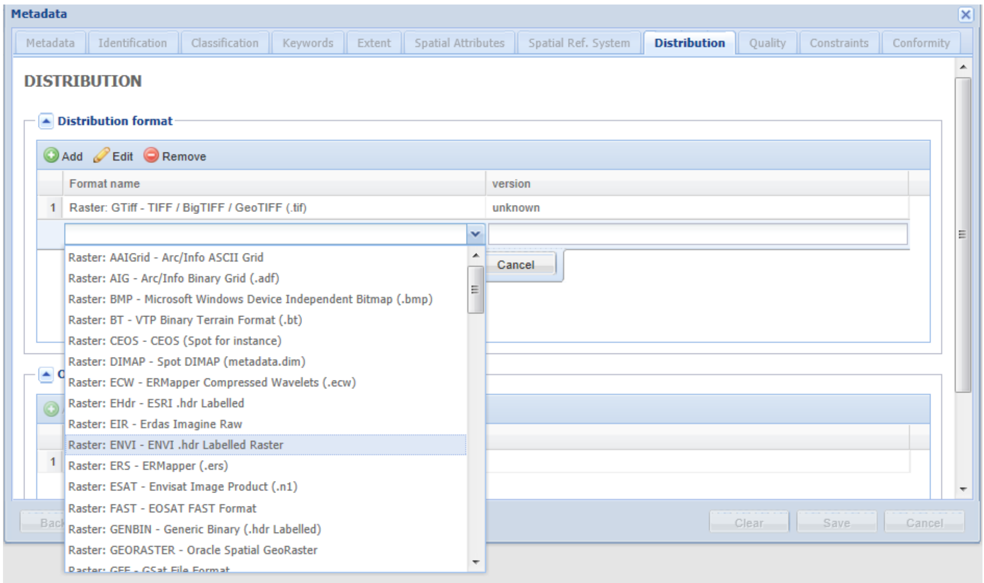


Figure 11 – Distribution format names on the metadata editor

8.2 Format Version *

Version of the format (date, number, etc.). The value domain of this metadata element is free text.

*8.3 Online resource: linkage **

Location (address) for on-line access using a Uniform Resource Locator (URL) address or similar addressing scheme. The resource locator defines the link(s) to the resource and/or the link to additional information about the resource. This metadata element is conditional for spatial dataset and spatial dataset serie; however it is mandatory if a URL is available to obtain more information on the resources and/or access related services. The value domain of this metadata element is a character string, commonly expressed as uniform resource locator (URL).

9. Quality

9.1 Lineage

This is a statement on process history and/or overall quality of the spatial data set (Figure 12). Where appropriate it may include a statement whether the data set has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity.

The value domain of this metadata element is free text.

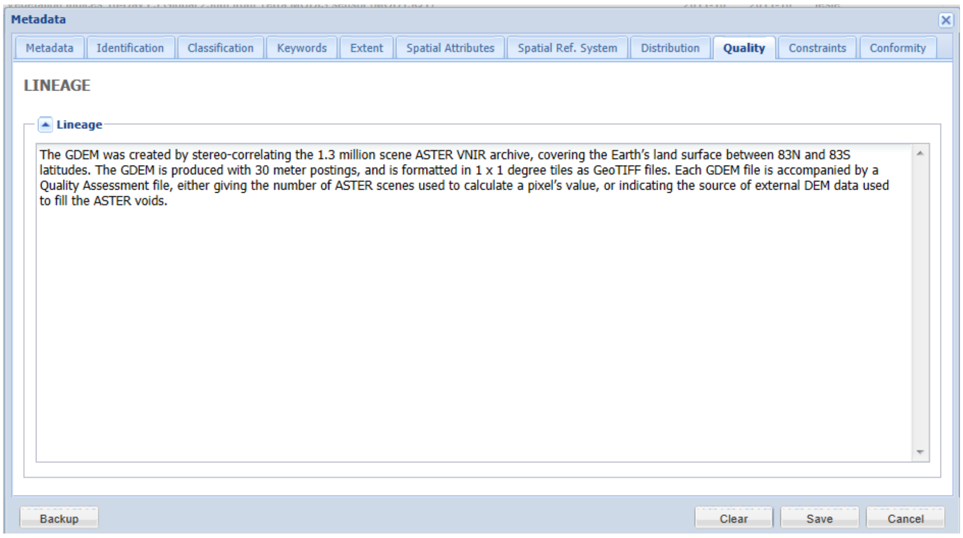


Figure 12 – Lineage information on the metadata editor

10. Constraint related to access and use

*10.1 Conditions applying to access and use **

This metadata element defines the conditions for access and use of spatial data sets and services, and where applicable, corresponding fees as required by Article 5(2)(b) and Article 11(2)(f) of Directive 2007/2/EC.

The value domain of this metadata element is free text but is advisable to use the suggestions on the metadata editor.

The element must have values. If no conditions apply to the access and use of the resource, "no conditions apply" shall be used. If conditions are unknown, "conditions unknown" shall be used. This element shall also provide information on any fees necessary to access and use the resource, if applicable, or refer to a uniform resource locator (URL) where information on fees is available.

*10.2 Limitations on public access **

When Member States limit public access to spatial data sets and spatial data services under Article 13 of Directive 2007/2/EC, this metadata element shall provide information on the limitations and the reasons for them.

If there are no limitations on public access, this metadata element shall indicate that fact.

The value domain of this metadata element is free text but is advisable to use the suggestions on the metadata editor.

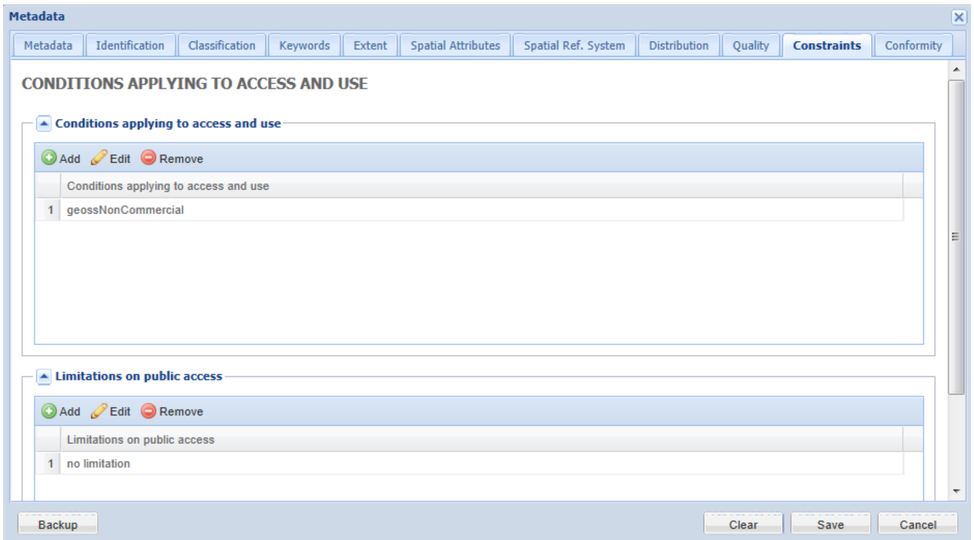


Figure 13 – Constraints to datasets access and use on the metadata editor

11. Conformity

11.1 Specifications

This is a citation of the implementing rules adopted under Article 7(1) of Directive 2007/2/EC (http://www.inspire-geoportal.eu/EUOSME_GEOPORTAL/userguide/eurlex_en.htm) or other specification to which a particular resource conforms (Figure 14).

A resource may conform to more than one implementing rules adopted under Article 7(1) of Directive 2007/2/EC or other specification.

11.1.1 Specification Title

This citation shall include the title of the implementing rules adopted under Article 7(1) of Directive 2007/2/EC or of the specification.

*11.1.2 Date **

This citation shall include the date (date of publication, date of last revision or of creation) of the implementing rules adopted under Article 7(1) of Directive 2007/2/EC or of the specification.

*11.1.3 Date type **

The citation of product specification shall include the data type of the implementing rules adopted under Article 7(1) of Directive 2007/2/EC or of the specification. The value domain of this metadata element a code list:

- Date of creation;
- Date of last revision;
- Date of publication.

*11.1.4 Degree **

This is the degree of conformity of the resource to the implementing rules adopted under Article 7(1) of Directive 2007/2/EC or other specification.

The value domain of this metadata element a code list:

- Not evaluated - Conformance has not been evaluated;
- Not conformant - The resource does not conform to the cited specification;
- Conformant - The resource is fully conformant with the cited specification.

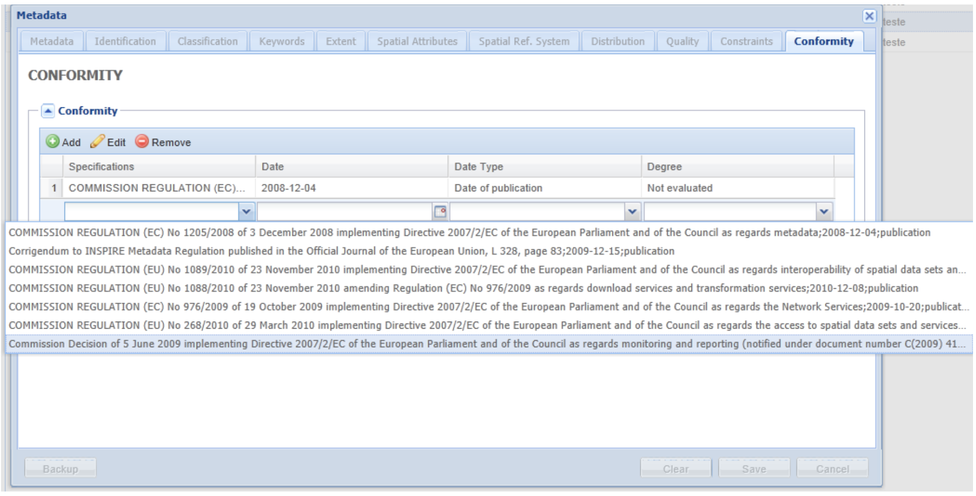
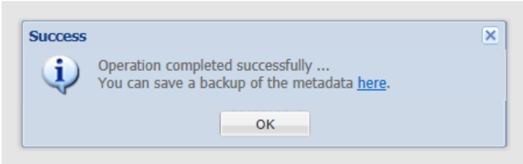


Figure 14 – Conformity information on the metadata editor

Final remarks

The metadata information introduced in the editor for each dataset (of the selected core datasets) must be saved and it is advisable to save a backup according to the steps indicated.


Success
Operation completed successfully ...
You can save a backup of the metadata here.
OK

Appendix II

Tables 1-5 present the core datasets identified by each Partner according to specific application contexts.

Table 1 – Core datasets identified per application context for UK site

Resource title	Resource identifier (metadata editor)
Application context: Sampling design for the collection of new on site campaigns for calibration or validation datasets of EODHaM habitat maps	
1 km grid squares	7bc0ddca-3e25-4f21-b256-c93ca05a729e
Land Cover Map 2000	55571dfd-ed67-4160-a0c0-f3516f48f7d1
Phase 1 Terr. AllWales.veg	bf4ee206-9276-4312-b2b7-cec7a5b6b694
Application context: Production of habitat maps within the specified quality requirements for BIO_SOS	
OS VM Railways Lines	e7756e17-a8db-4acb-9926-bde7fd2974dd
OS VM Roads	928972ef-5e42-44aa-a8ee-9a2f6460ea3e
EA - Welsh Reservoirs	b330fce2-bf1a-4fc0-8bbb-792a3ea48463
Mean High Water Mark	74b18d44-ba20-4ffb-a5f1-70dada9481a0
LPIS Field Information	60a548a4-6783-4a6e-bb09-fd67990636f6
Air photos 2006 Cowi-Vexcel	0ea91e09-0ed7-4186-9f8e-ece3e470eb9b
Hyperspectral (EAGLE) Cors Fochno	9b595f76-a0a6-4a77-baf0-48d7b6b060f7
LiDAR data, Cors Fochno	2c441eda-a40e-4fc3-b155-66b7517e26b3
Contours 5m SN	3ec187b4-4589-46df-82e1-35f7f48a59e5
Application context: Evaluation of the effects of specific pressures on habitats and biodiversity	
Protected Sites: National Nature Reserves - Final Bounds - GIS Dataset	C501F266-E660-0001-B742-1C60B5101988
Sites (SAC)	19d8c2de-c77e-42bd-84e2-cf652ec2de6c
SSSI Qualifying features	6b8625fd-f80c-4637-bf13-a589007de2bc

Table 2 – Core datasets identified per application context for Greece test sites

Resource title	Resource identifier (metadata editor)
Application context: Sampling design for the collection of new on site campaigns to collect new data on pressures	
CLC 1990 core	C5032F81-4EE0-0001-C2E8-1E01696FE6D0
CLC2000 core	C50331A3-E510-0001-2EC5-12E223261453
Natura 2000 network SPA	C50332C4-9B40-0001-F3CA-7E9E1268155B
Orthoimagery GRD 50cm core	C50339AE-31C0-0001-F318-25A022801FC2
EO data (Landsat USGS) core	C5033783-3400-0001-5212-B50011E01280
Species of Annex II of Directive 92/43/CE distribution maps	-
Phytosociological relevés (samples)taken in 1999-2000	-
Application context: Sampling design for the collection of new on site campaigns to collect new data on biodiversity	
CLC2000 core	C50331A3-E510-0001-2EC5-12E223261453
Natura 2000 network SPA	C50332C4-9B40-0001-F3CA-7E9E1268155B
Natura 2000 network SCI	C50332C4-9B40-0001-F3CA-7E9E1268155B
Orthoimagery GRD 50cm core	C50339AE-31C0-0001-F318-25A022801FC2
EO data (Landsat USGS) core	C5033783-3400-0001-5212-B50011E01280

Species of Annex II of Directive 92/43/CE distribution maps	-
Phytosociological relevés (samples)taken in 1999-2000	-

Table 2 – Core datasets identified per application context for Greece test sites (cont.)

Application context: Sampling design for the collection of new on site campaigns for calibration or validation datasets of EODHaM habitat maps	
ASTER Global Digital Elevation Model (GDEM) core	C50338A6-9050-0001-F080-33B02FC61265
CLC2000 core	C50331A3-E510-0001-2EC5-12E223261453
Natura 2000 network SPA	C50332C4-9B40-0001-F3CA-7E9E1268155B
Natura 2000 network SCI	C50332C4-9B40-0001-F3CA-7E9E1268155B
Orthoimagery GRD 50cm core	C50339AE-31C0-0001-F318-25A022801FC2
EO data (Landsat USGS) core	C5033783-3400-0001-5212-B50011E01280
Phytosociological relevés (samples)taken in 1999-2000	-
Application context: Evaluation of the effects of specific pressures on habitats and biodiversity	
ASTER Global Digital Elevation Model (GDEM) core	C50338A6-9050-0001-F080-33B02FC61265
CLC 1990 core	C5032F81-4EE0-0001-C2E8-1E01696FE6D0
CLC2000 core	C50331A3-E510-0001-2EC5-12E223261453
Natura 2000 network SPA	C50332C4-9B40-0001-F3CA-7E9E1268155B
Natura 2000 network SCI	C50332C4-9B40-0001-F3CA-7E9E1268155B
Orthoimagery GRD 50cm core	C50339AE-31C0-0001-F318-25A022801FC2
EO data (Landsat USGS) core	C5033783-3400-0001-5212-B50011E01280
Species of Annex II of Directive 92/43/CE distribution maps	-
Phytosociological relevés (samples)taken in 1999-2000	-
Application context: Production of habitat maps within the specified quality requirements for BIO_SOS	
ASTER Global Digital Elevation Model (GDEM) core	C50338A6-9050-0001-F080-33B02FC61265
Natura 2000 network SPA	C50332C4-9B40-0001-F3CA-7E9E1268155B
Natura 2000 network SCI	C50332C4-9B40-0001-F3CA-7E9E1268155B
Phytosociological relevés (samples)taken in 1999-2000	-

Table 3 – Core datasets identified per application context for Netherland site

Resource title	Resource identifier (metadata editor)
Application context: Sampling design for the collection of new on site campaigns for calibration or validation datasets of EODHaM habitat maps	
National Land Use Netherlands - LGN6	fbf5cda7-f1e9-4834-988b-963c02c66fb0
Natura 2000 areas as of January 2006	db4e1245-51a1-45a4-bd7b-9bff5b440196
Aerial photography of Eder and Ginkelse Heathland	-
Application context: Production of habitat maps within the specified quality requirements for BIO_SOS	
AHN 2000 - 5m grid (Digital Elevation model)	c8d58c34-c200-48a6-94a7-be793e56d9a5
Basic Mapping of Nature 2004	dde8b7fc-fad2-4d6d-ad61-87e595c638f1
Soil map of the Netherlands, scale 1:50000 with peat mapping, version 2006	683e1313-9a57-4670-bfc4-4a302b3ac38d
Structure mapping of Edese and Ginkelse Heathland 2003	3f441db5-c2cb-4a6b-b841-084ed0a14dc4
Vegetation mapping of Edese and Ginkelse Heathland and Wekeromse Sand 2009	0b75516d-2153-4065-8982-6f9ed230279b
BRT TOP10NL	C503CE77-6D20-0001-A0FC-1CE0F5201CAB
Application context: Evaluation of the effects of specific pressures on habitats and biodiversity	
National Vegetation Releves Database - LVD	C503E401-71A0-0001-5F68-19D71671122B

Table 4 – Core datasets identified per application context for Brasil site

Resource title	Resource identifier (metadata editor)
Application context: not specified	
SPOT5	d89c80af-d3ca-494c-997a-2f533e159bd3
MODIS Vegetation index MOD13Q1 250 m h12v09	b66b7edc-419c-4b58-b92d-b7d692220850
ASTER GDEM	24fb0851-8f80-4003-80cc-79e3c2b0649b
LANDSAT	ee096675-b17d-4e6a-9c2e-8671ba7d45d6
BR_indigenous_lands	1d5be2a4-fcb3-4d84-a185-85bb0807f601
BR_admin	ab507814-5e4c-4996-9bbd-9f1d5c8a6e19
BR_conservation_units	718a6a3d-f987-4036-a50c-b1b52cf14491
AMZ_hydrography	c042667e-8395-4bdd-872b-74ca435ab4e4
AMZ_PRODES	de69a59f-630a-4d87-823a-a6af9d2b1fa7

Table 5 – Core datasets identified per application context for Portugal sites

Resource title	Resource identifier (metadata editor)
Application context: Sampling design for the collection of new on site campaigns to collect new data on pressures	
CLC 1990 [SP]	b9552ab3-fb93-4696-9fa7-00be8dee417c
CLC 2000 [SP]	d6c286ae-3c17-4004-a7c7-e05241517d8b
CLC 2006 [SP]	1175be03-fbb6-4b7d-868a-f11eca127b1c
ASTER Global Digital Elevation Model (GDEM) [SP]	c56f60ec-f93a-4e5c-9475-2e35b1eff74d
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	8b9db7b2-0c2c-4205-a510-568abd91e719
COS 1990 [S]	bc29937f-591a-4af6-8ea1-a834fb57fc08
COS 1990 [P]	8c2811c9-baee-490b-98aa-eb22abe2ed2b
COS 2006 [P]	3975d555-a175-4ae4-b66c-4eb721be12e9
Contour_10m [P]	eb2155f5-ead2-4925-a496-941bee09ab12
Elevation points [P]	5933069c-0dcd-498c-86ac-d30b67852ff4
Geodesic vertex [P]	a46746ce-40bd-4903-ad39-2d1d6f1589cf
Phytosociological associations / vegetation types distribution [P]	51861ea4-91c2-41ae-8996-ef8ad9d46fec
Plant species records / distribution data for mainland Portugal from phytoso-ciological inventories [SP]	42fe1f5e-5884-41cf-94e3-d90bf1587753
Application context: Sampling design for the collection of new on site campaigns to collect new data on biodiversity	
CLC 1990 [SP]	b9552ab3-fb93-4696-9fa7-00be8dee417c
CLC 2000 [SP]	d6c286ae-3c17-4004-a7c7-e05241517d8b
CLC 2006 [SP]	1175be03-fbb6-4b7d-868a-f11eca127b1c
ASTER Global Digital Elevation Model (GDEM) [SP]	c56f60ec-f93a-4e5c-9475-2e35b1eff74d
Vegetation Indices 16-Day L3 Global 250m from Terra MODIS sensor (MOD13Q1) [SP]	8b9db7b2-0c2c-4205-a510-568abd91e719
COS 1990 [S]	bc29937f-591a-4af6-8ea1-a834fb57fc08
COS 1990 [P]	8c2811c9-baee-490b-98aa-eb22abe2ed2b
COS 2006 [P]	3975d555-a175-4ae4-b66c-4eb721be12e9
Contour_10m [P]	eb2155f5-ead2-4925-a496-941bee09ab12
Elevation points [P]	5933069c-0dcd-498c-86ac-d30b67852ff4
Geodesic vertex [P]	a46746ce-40bd-4903-ad39-2d1d6f1589cf
Phytosociological associations / vegetation types distribution [P]	51861ea4-91c2-41ae-8996-ef8ad9d46fec
Plant species records / distribution data for mainland Portugal from phytoso-ciological inventories [SP]	42fe1f5e-5884-41cf-94e3-d90bf1587753

Application context: Sampling design for the collection of new on site campaigns for calibration or validation datasets of EODHaM habitat maps	
CLC 1990 [SP]	b9552ab3-fb93-4696-9fa7-00be8dee417c
CLC 2000 [SP]	d6c286ae-3c17-4004-a7c7-e05241517d8b
CLC 2006 [SP]	1175be03-fbb6-4b7d-868a-f11eca127b1c
COS 1990 [S]	bc29937f-591a-4af6-8ea1-a834fb57fc08
COS 1990 [P]	8c2811c9-baee-490b-98aa-eb22abe2ed2b
COS 2006 [P]	3975d555-a175-4ae4-b66c-4eb721be12e9
Application context: Evaluation of the effects of specific pressures on habitats and biodiversity	
COS 1990 [S]	bc29937f-591a-4af6-8ea1-a834fb57fc08
COS 1990 [P]	8c2811c9-baee-490b-98aa-eb22abe2ed2b
COS 2006 [P]	3975d555-a175-4ae4-b66c-4eb721be12e9
Phytosociological associations / vegetation types distribution [P]	51861ea4-91c2-41ae-8996-ef8ad9d46fec
Plant species records / distribution data for mainland Portugal from phytoso-ciological inventories [SP]	42fe1f5e-5884-41cf-94e3-d90bf1587753
Application context: Production of habitat maps within the specified quality requirements for BIO_SOS	
ASTER Global Digital Elevation Model (GDEM) [SP]	c56f60ec-f93a-4e5c-9475-2e35b1eff74d
National Road Plan 2000 [SP]	fe8becf9-d21e-4fd5-863f-413c3707c66f
Main Road Network [P]	4446647d-e99a-4234-825e-f2370dff299c
Contour_10m [P]	eb2155f5-ead2-4925-a496-941bee09ab12
Elevation points [P]	5933069c-0dcd-498c-86ac-d30b67852ff4
Geodesic vertex [P]	a46746ce-40bd-4903-ad39-2d1d6f1589cf
EO data from Landsat5 TM sensor_20030711 [P]	84e0e2df-3a95-4f89-8ae5-96a603d8ea60
EO data from Landsat5 TM sensor_20070706 [P]	20ccd36a-7784-4c4d-b5fa-31d31301581b
EO data from Landsat5 TM sensor_20100425 [P]	11b7f8ef-b226-4323-8cb1-03049f4adae7
EO data from Landsat5 TM sensor_20100730 [P]	dc29acdc-bbc8-4b80-aa03-bc5f41c8e060
EO data from Landsat7 ETM+ sensor_19990708 [P]	8e908d09-7b45-4adb-9ac0-d70cdcdbc854
EO data from Landsat7 ETM+ sensor_20010526 [P]	6288497f-59ed-43ab-8148-b0cb19861c4b
EO data from Landsat7 ETM+ sensor_20010915 [P]	c89bb70b-f6e6-4c14-ac14-e54a33fe5445
EO data from Landsat5 TM sensor_19900902 [S]	36fe2376-7918-44fc-aba4-f91bbe08c6a7
EO data from Landsat4 TM sensor_19910524 [S]	d03cef63-0a45-4278-b1fe-6dbce8828e56
EO data from Landsat5 TM sensor_20030805 [S]	ae6fb0af-9329-47bd-b70e-5830ac5c6d51
EO data from Landsat5 TM sensor_20031008 [S]	7d4fe522-e6d7-48a5-94a4-e7e2e245e743
EO data from Landsat5 TM sensor_20070325 [S]	b2261a79-dc9b-4efb-81f8-d557378f9c1e
EO data from Landsat5 TM sensor_20070731 [S]	05b74e29-5a06-44fe-9bc0-93fa3abaeb9d
EO data from Landsat5 TM sensor_20090805 [S]	b5268f5e-0957-4c6d-870d-4d17f516622c
EO data from Landsat5 TM sensor_20100520 [S]	376f8947-f0ca-4051-8400-dc69d12d7415
EO data from Landsat5 TM sensor_20100707 [S]	87a23642-19d8-4e29-85b8-b350d521c610
EO data from Landsat7 ETM+ sensor_20000804 [S]	28070f7c-ba61-4770-b611-22899b41a71e
EO data from Landsat7 ETM+ sensor_20030407 [S]	150e5916-a29b-47c4-83b5-79b057837410

Appendix III

Prototype examples of graphical user interfaces for external quality evaluation

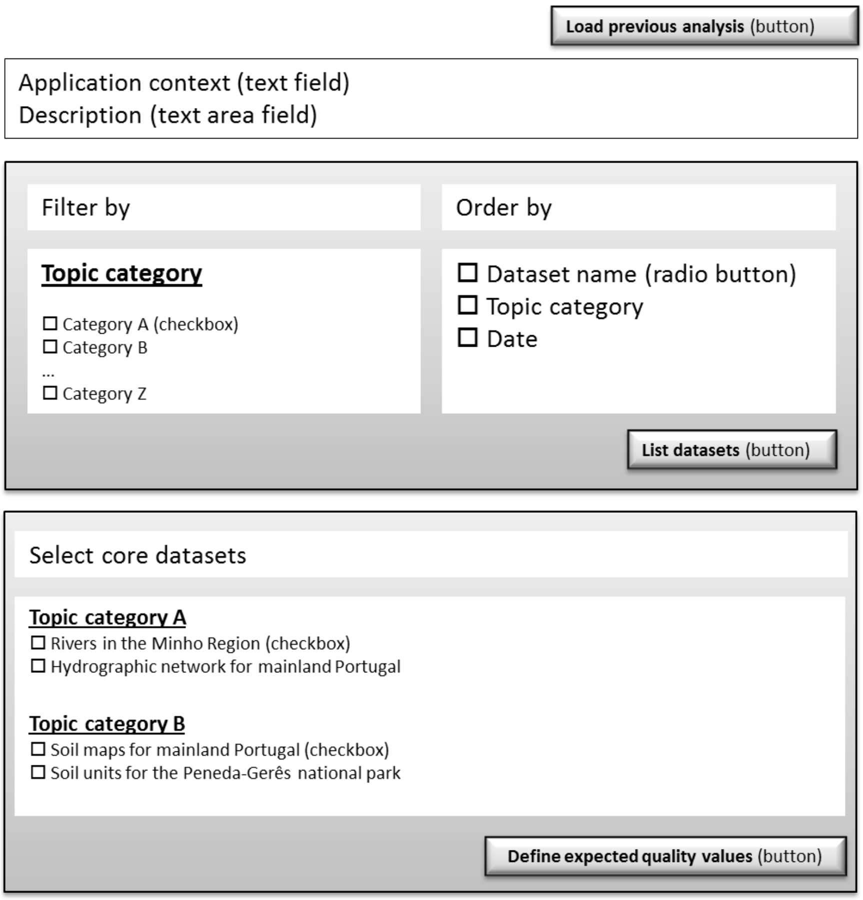
The external quality evaluation (EQE) methodology is already structured for integration in the WebGIS platform and will be integrated soon, interconnecting with the metadata editor and with search tools. In this Appendix are presented some prototype examples of the graphical user interfaces to be used for the EQE within the platform.

viii) Graphical interfaces by EQE objective

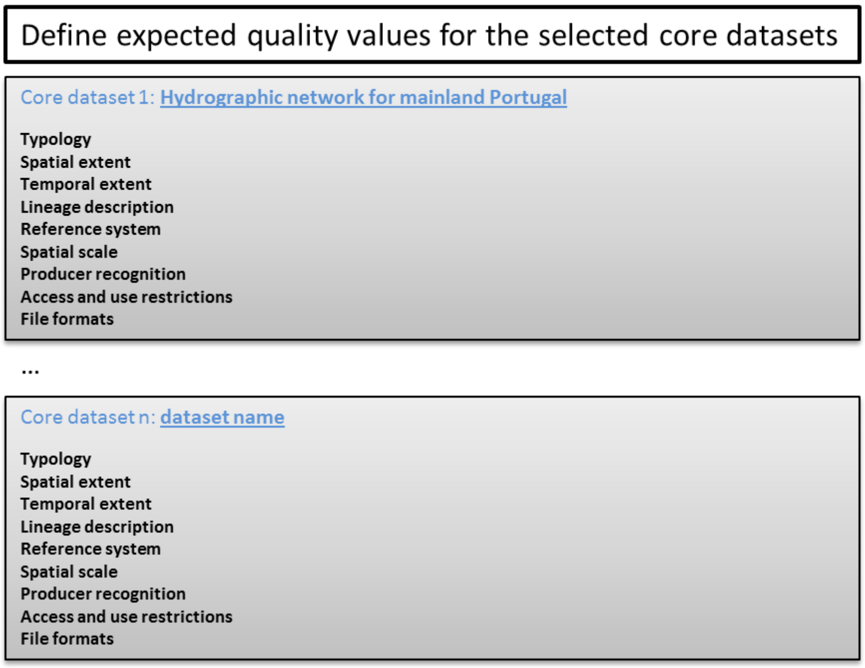
As previously mentioned EQE may be used for two different purposes (A1 and A2) which affect the way the application will behave. In order to facilitate the description of the application GUI, a minimum suggested set of elements that should be displayed is described in the following sections.

A1 – Data search, querying and ranking using EQE concepts

[A1-I1]



[A1-I2]



[A1-I2a] (detail on UI box element)

Core dataset 1: **Hydrographic network for Portugal** | See metadata (link or button)

Typology | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Topic categories (dropdown list)

Spatial extent | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
North Bound Latitude (text field)
East Bound Longitude (text field)
South Bound Latitude (text field)
West Bound Longitude (text field)
Draw bounding box (map widget)
Percentage cover (slider bar)

Temporal extent | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Start date (text field/calendar widget)
End date (text field/calendar widget)
Percentage cover (slider bar)

Lineage description | Is CF? (checkbox)
Lineage description (checkbox)

Reference system | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Reference system (dropdown list)

Spatial scale | + Add equivalent scale (button) | + Add resolution distance (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Admissible equivalent scale range (slider bar with 2 elements/minimum and maximum value)
Admissible resolution distance range (slider bar with 2 elements/minimum and maximum value) | Measurement units (dropdown list)

Producer recognition | Is CF? (checkbox)
Producer recognition (dropdown list)

Access and use restrictions | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Conditions applying to access and use (dropdown list)

File formats | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Distribution format (dropdown list)

[A1-I3]

External quality assessment summary

Mean fitness value for application context: (between 0 and 100%)

Core dataset 1: **Hydrographic network for mainland Portugal**

Diagnosis (fit, partially fit, unfit)
Fitness value (between 0 and 100%)
Critical indicators in fit condition (if any, a list of QI that were considered fit)
Critical indicators in partial fit condition (if any, a list of QI that were considered unfit but were not regarded as CFs)
Critical indicators in unfit condition (if any, a list of QI that were considered unfit but were defined as CFs)

…

Core dataset n: **dataset name**

Diagnosis (fit, partially fit, unfit)
Fitness value (between 0 and 100%)
Critical indicators in fit condition
Critical indicators in partial fit condition
Critical indicators in unfit condition

Save analysis (button) | **Print summary** (button) | **Export summary** (button)

A2 – EQE as an assessment procedure or diagnostic tool

[A2-I1]

Load previous search parameters (button)

Filter by

☐ Topic category (checkbox) | + Add value (button) | Value (dropdown list)
☐ Keyword (checkbox) | + Add value (button) | [Metadata field (dropdown list) | Value (text field)]

Define search parameters

Typology | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Topic categories (dropdown list)

Spatial extent | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
North Bound Latitude (text field) | East Bound Longitude (text field) | South Bound Latitude (text field) | West Bound Longitude (text field)
Draw bounding box (map widget)
Percentage cover (slider bar)

Temporal extent | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Start date (text field/calendar widget)
End date (text field/calendar widget)
Percentage cover (slider bar)

Lineage description | Is CF? (checkbox)
Lineage description (checkbox)

Reference system | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Reference system (dropdown list)

Spatial scale | + Add equivalent scale (button) | + Add resolution distance (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Admissible equivalent scale range (slider bar with 2 elements/minimum and maximum value)
Admissible resolution distance range (slider bar with 2 elements/minimum and maximum value) | Measurement units (dropdown list)

Producer recognition | Is CF? (checkbox)
Producer recognition (dropdown list)

Access and use restrictions | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Conditions applying to access and use (dropdown list)

File formats | + Add value (button) | Match all/any (dropdown list) | Is CF? (checkbox)
Distribution format (dropdown list)

Search (button)

[A2-I2]

Search results

Legend

Fit	Partially fit	Unfit
✓	❙	✗

1. **Fit datasets**

Datasets	Fitness value	Quality indicators status								
		Typology	Spatial extent	Temporal extent	Lineage description	Reference system	Spatial scale	Producer recognition	Access and use restrictions	File formats
Dataset name 1	100%	✓	✓	✓	✓	✓	✓	✓	✓	✓
...	100%	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dataset name n	100%	✓	✓	✓	✓	✓	✓	✓	✓	✓

2. Partially fit datasets

Datasets	Fitness value	Quality indicators status								
		Typology	Spatial extent	Temporal extent	Lineage description	Reference system	Spatial scale	Producer recognition	Access and use restrictions	File formats
Dataset name 1	0.89%	✓	✓	▮	✓	✓	✓	✓	✓	✓
...	0.78%	✓	✓	✓	▮	▮	✓	✓	✓	✓
Dataset name n	0.56%	✓	▮	▮	▮	▮	✓	✓	✓	✓

3. Unfit datasets

Datasets	Fitness value	Quality indicators status								
		Typology	Spatial extent	Temporal extent	Lineage description	Reference system	Spatial scale	Producer recognition	Access and use restrictions	File formats
Dataset name 1	0.89%	✓	✓	✗	✓	✓	✓	✓	✓	✓
...	0.78%	✓	✓	✓	✗	▮	✓	✓	✓	✓
Dataset name n	0.56%	✓	▮	✗	✗	▮	✓	✓	✓	✓

Save search parameters (button) **Print summary** (button) **Export summary** (button)

Table A2.1 – Expected Quality matrix filled in by Partners for each application context

Appendix IV

Identification of the application context:																						
Resource title (1)	File identifier (2)	Quality indicators (3)																				
		Typology	CF (4) (Y=1;N=0)	Spatial extent	CF (4) (Y=1;N=0)	% cover (5)	Temporal extent	CF (4) (Y=1;N=0)	% cover (5)	Lineage description (Y=1; N=0)	CF (4) (Y=1;N=0)	Reference system	CF (4) (Y=1;N=0)	Spatial scale		CF (4) (Y=1;N=0)	Producer recognition	CF (4) (Y=1;N=0)	Acess and use restrictions	CF (4) (Y=1;N=0)	File formats	CF (4) (Y=1;N=0)
														Equivalent Scale	Resolution Distance							
Expected quality matrix (5)																						

Table A2.2 –Internal Quality matrix filled in by Partners for each application context

Identification of the application context:												
Resource title (1)	File identifier (2)	Quality indicators (3)										
		Typology	Spatial extent	Temporal extent	Lineage description	Reference system	Spatial scale			Producer recognition	Acess and use restrictions	File formats
							Equivalent Scale	Spatial Resolution	Unit of measure			
Internal quality matrix (5)												

10. References

Abugessaisa, I. and Östman; A. (2011).Testing-SDI: E-government Prospective, Requirements, and Challenges International Journal of Public Information Systems Vol. 2011:1, pp. 11-47.

Bédard, Y. and D. Vallière (1995). Qualité des données à référence spatiale dans un contexte gouvernemental. Technical report. Laval University, Canada

Devillers, R., Bédard, Y., Jeansoulin, R., Moulin, B. (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. International Journal of Geographical Information Science 21(3): 261-282.

EC, 2009. INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, v. 1.1. European Commission (EC), Joint Research Centre, http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf

Federal Geographic Data Committee (2000). Content Standard for Digital Geospatial Metadata Workbook Version 2.0. Federal Geographic Data Committee. Washington, D.C.

Frank, A.U., Grum, E., Vasseur, B., 2004. Procedures to select the best dataset for a task. In: M.J. Egenhofer, H. Miller, C. Freksa (Eds). Proceedings of the Third International Conference on Geographic Information Science. University of Maryland, USA, pp: 20-23

GeoNetwork opensource, 2011a. User Manual v2.6.4, Open Source Geospatial Foundation, http://geonetwork-opensource.org/manuals/2.6.4/users/index.html, accessed on 23/11/2011.

GeoNetwork opensource, 2011b. Developer Manual v2.6.4, Open Source Geospatial Foundation, http://geonetwork-opensource.org/manuals/2.6.4/developer/index.html , accessed on 23/11/2011.

GeoServer, 2011. GeoServer User Manual, GeoServer, http://docs.geoserver.org/stable/en/user/, accessed on 23/11/2011.

Gyamfi-Aidoo, J.; Schwabe, C.; Govender, S. (2006). Determination of the fundamental Geospatial datasets for Africa through a user needs analysis – A synthesis report. In: Determination of the fundamental Geospatial datasets for Africa through a user needs analysis. Human Sciences Research Council (HSRC) and EIS AFRICA

INSPIRE (2008). D3.7.2 Draft Implementing Rules for View Services. At http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3.7.2_Draft_IR_View_Services_v3.0.pdf

ISO 19115:2005 Geographic Information - Metadata.

ISO 9241-11:1998; Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability

ISO 9241-210:2010; Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems

ISO, 2007. ISO/TS 19139:2007, Geographic information - Metadata - XML schema implementation.

ISO/PAS 20282-3:2007; Ease of operation of everyday products -- Part 3: Test method for consumer products

Lopez-Pellicer F.J., Béjar R., Florczyk A.J., Muro-Medrano P.R., Zarazaga-Soria F.J., 2011. A review of the implementation of OGC Web Services across Europe, International Journal of Spatial Data Infrastructures Research, 2011, Vol.6, 168-186

Markov Z., Larose D. (2007). *Data Mining the Web - Uncovering Patterns in Web Content, Structure, and Usage*, New York: Wiley, ISBN 978-0-471-66655-4.213 stran.

Nerbert, D.D. (2004). Developing Spatial Data Infrastructures: The SDI Cookbook. (GSDI Cookbook, Version 2.0). Global Spatial Data Infrastructure. pp. 171

Nielsen, J. 2000, *Designing Web Usability*, New Riders Publishing, Indianapolis

OGC, 2007. OpenGIS® Catalogue Services Specification, version 2.0.2, Open Geospatial Consortium Inc.

OGC, 2010. OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 2: SQL option, version 1.2.1, Open Geospatial Consortium Inc.

PGDG, 2011. PostgreSQL 9.1.1 Documentation, The PostgreSQL Global Development Group, http://www.postgresql.org/docs/9.1/static/index.html, accessed on 23/11/2012.

PostGIS, 2011. PostGIS, Refractions Research, http://postgis.refractions.net/, accessed on 23/11/2012.

Rajabifard A., Kalantari M., Binns A., 2009. SDI and Metadata Entry and Updating Tools, SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors). Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48.

RDM Working Group (2002). Reference Data and Metadata Position Paper. INSPIRE Infrastructure for Spatial Information in Europe (Eds.: Daniel Rase, Anette Björnsson, Mark Probert, Marie-France Haupt). EUROSTAT

SDI Cookbook (2009). Spatial Data Infrastructure Cookbook. 150 pp. http://www.gsdi.org/gsdicookbookindex

Starbuck, M. (2001). Review of availability and accessibility of geospatial data in the greater Mekong subregion. A consultancy report in support of the Subregional Environmental Monitoring and Information Systems II Project (SEMIS II). United States Geological Survey

Vasseur, B., Jeansoulin, R., Devillers, R., Frank, A. 2006. External data quality of geographical applications: an ontological approach. In: Devillers, R., Jeansoulin, R. (Eds). Fundamentals of Spatial Data Quality, 2006. ISTE – GIS Series, pp: 31-42