Project Title:	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Project number:	641762
Project Acronym:	ECOPOTENTIAL
Proposal full title:	IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Type:	Research and innovation actions
Work program topics addressed:	SC5-16-2014: "Making Earth Observation and Monitoring Data usable for ecosystem modelling and services"

D5.3 Framework for user-oriented quality evaluation routines

Due date of deliverable:	30th November 2017
Actual submission date:	13th December 217
Version:	v1
Main Authors:	Joaquim Alonso (ICETA), Pedro Castro (ICETA), Ivone Martins (ICETA), João P. Honrado (ICETA), Miguel Calafate (ICETA), João Gonçalves (ICETA), Johannes Peterseil (EAA), António Monteiro (ICETA), Cláudia Santos (ICETA), Salvador Arenas-Castro (ICETA), Bruno Marcos (ICETA), Dimitris Poursanidis (FORTH), Jordi Prados (STARLAB), Chiara Pratola (STARLAB).

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Project ref. number	641762
Project title	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS

Deliverable title	Framework for user-oriented quality evaluation routines
Deliverable number	D5.3
Deliverable version	1
Contractual date of delivery	30.11.2017
Actual date of delivery	13.12.2017
Document status	Final
Document version	1
Online access	www.ecopotential-project.eu
Diffusion	Public
Nature of deliverable	Other
Work package	WP 5
Partner responsible	13 – Instituto de Ciências, Tecnologias e Agroambiente (ICETA) CIBIO/InBIO, Universidade do Porto, Portugal
Author(s)	Joaquim Alonso, Pedro Castro, Ivone Martins, João P. Honrado, Miguel Calafate, João Gonçalves, Johannes Peterseil, António Monteiro, Cláudia Santos, Salvador Arenas-Castro, Bruno Marcos, Dimitris Poursanidis, Jordi Prados, Chiara Pratola.
Editor	Joaquim Alonso
Approved by	
EC Project Officer	Gaëlle Le Bouler

Abstract	Knowledge collaborative networks of biodiversity analysis and ecosystems services modelling imply spatial data management and spatial data quality assessment (SDQA). SDQA improvements require conceptual proposals and spatial data standards advances namely in metadata profiles and data users capacity building as well as, development of spatial data quality assessment tools and data quality management routines.
Keywords	Spatial metadata, ISO 191157; Biodiversity, Ecosystem services quality assessment, quality management, ThemIsE, collaborative science;

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Table of Contents

List of tables

List of figures

Terms and abbreviations

ABCD	Access to Biological Collection Data
AC	Atmospheric conditions
AD	Addresses
AF	Agricultural and aquaculture facilities
AM	Area management/ restriction/ regulation zones and reporting units
API	Application Programming Interface
ARATOS	Aratos Technologies S.A.
AU	Administrative units
AVDC	Aura Validation Data Centre
BGU	Ben-Gurion University of the Negev
BIO_SOS	Biodiversity Multi-SOurce Monitoring System - FP7 project
BISE	Biodiversity Information System for Europe
BR	Bio-geographical regions
BU	Buildings
CERTH	Centre for Research and Technology Hellas
CESBIO-UPS	Center for the Study of the Biosphere from Space, Université Paul Sabatier Toulouse III
CIBIO/InBIO	Research Center in Biodiversity and Genetic Resources/ InBIO Associate Laboratory
CNR	Consiglio Nazionale delle Ricerche
CNRS-UMR	Centre National de la Recherche Scientifique, MARine Biodiversity, Exploitation and Conservation
COPERNICUS	Copernicus: Europe's eyes on Earth
COBWEB	COBWEB: Citizen Observatory Web
CP	Cadastral parcels
CREAF	Center for Research in Ecology and Forestry Applications
CSIC	Estación Biológica de Doñana
CSIR	Council for Scientific and Industrial Research
CSS	Cascading Style Sheets
CSW	Catalogue Services for the Web

D2.2	Deliverable 2.2 “EO-driven Essential Variables”
D5.2	Deliverable 5.2 “Metadata for pre-existing datasets”
D5.3	Deliverable 5.3 “Framework for user-oriented quality evaluation routines”
D5.7	Deliverable 5.7 “Database of preexisting and new data”
DC	Dublin Core metadata
DCMI	Dublin Core Metadata Initiative
DEIMS	Dynamic Ecological Information Management System
DEIMS-SDR Dataset MD	Dynamic Ecological Information Management System - Site and Dataset Registry, Dataset MD Model
DELTARES	Stichting Deltares
DLR	German Aerospace Center
DQ	Data quality
DQM	Data quality measures
DwC	Darwin Core metadata
EAA	Umweltbundesamt GmbH
EC	European Commission
ECOP	ECOPOTENTIAL project
ECOPOTENTIAL	Improving future ecosystem benefits through earth observations
EF	Environmental monitoring facilities
EL	Elevation
EM	Ecological Modelling
EML	Ecological Modelling Framework
EnvThes	Environmental Thesaurus
EO	Earth Observation
EOS	Earth Observing System
EPFL	École Polytechnique Fédérale de Lausanne
EQDaM	External Quality of [Spatial] Data [from] Metadata
ER	Energy resources
ESA	European Space Agency
ESDIN	European Spatial Data Infrastructure with a Best Practice Network

ESL	Environment Systems Limited
ETH	Swiss Federal Institute of Technology Zurich
EU	European Union
EU BON	European Biodiversity Observation Network
EUNIS	European Nature Information System
EURAC	Accademia Europea di Bolzano
EVDC	Atmospheric Validation Data Centre
EVs	Essential Variables
FORTH	Foundation for Research and Technology Hellas
GALILEO	Galileo is Europe's own global navigation satellite system
GBIF	Global Biodiversity Information Facility
GCI	GEOSS Common Infrastructure
GCO	Geographic Common extensible markup language
GE	Geology
GECA	Generic Embodied Conversational Agent
GeoViQua	QUAlity aware VIsualization for the Global Earth Observation System of systems Project
GEOMS	Generic Earth Observation Metadata Standard
GEOSS	Global Earth Observation System of Systems
GG	Geographical grid systems
GIS	Geographic Information Systems
GMD	Geographic Metadata extensible markup language
GMES	Global Monitoring for Environment and Security
GMX	Geographic Metadata XML Schema
GN	Geographical names
GSDI	Global Spatial Data Infrastructure
GSR	Geographic Spatial Referencing extensible markup language
GSS	Geographic Spatial Schema extensible markup language
GTS	Geographic Temporal Schema extensible markup language
GUF	Geospatial User Feedback
GUI	Graphical User Interface

GUID	Globally Unique Identifiers
HB	Habitats and biotopes
HDF	Hierarchical Data Format
HH	Human health and safety
HIO	PSI Hydrobiological Institute-OHRID
HTML	HyperText Markup Language
HY	Hydrography
ICETA	Instituto de Ciências, Tecnologias e Agroambiente, Universidade do Porto
CIBIO/InBIO	Research Center in Biodiversity and Genetic Resources/ InBIO Associate Laboratory
iDiv-MLU	Martin Luther University Halle-Wittenberg
IETF	Internet Engineering Task Force
IISTA-UGR	Andalusian Institute for Earth System Research, Universidad de Granada
IND_CHANGE	INDicator-based modelling tools to predict landscape CHANGE and to improve the application of social-ecological research in adaptive land management
INPA	Israel Nature and Parks Authority
INSPIRE	Infrastructure for Spatial Information in the European Community
INSPIRE Directive	Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007
IoT	Internet of things
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
ISO	International Organization for Standardization
ISO/DIS	ISO Draft International Standard
ISO/IEC	ISO International Electrotechnical Commission
ISO/TC	ISO Technical Committees
ISO/TS	ISO Technical Specification
ISPRA	Institute for Environmental Protection and Research
IST	Instituto Superior Técnico, Universidade de Lisboa
JS	JavaScript
JSON	JavaScript Object Notation
KIT	Karlsruhe Institute of Technology
KNB	Knowledge network for Biocomplexity

KU	Klaipeda University
LC	Land cover
LifeWatch	E-Science European Infrastructure for Biodiversity and Ecosystem Research
LSE	London School of Economics and Political Science
LTER	Long-Term Ecological Research Network
LU	Land use
MD	Metadata
MF	Meteorological geographical features
MfN	Museum für Naturkunde, Berlin
MR	Mineral resources
NatureSDIplus	Best Practice Network for SDI in nature conservation
NDACC	Network for Detection of Atmospheric Composition Change
NextGEOSS	The next generation of the Global Earth Observation System of Systems
NIOZ	Royal Netherlands Institute for Sea Research
NZ	Natural risk zones
OECD	Organisation for Economic Co-operation and Development
OF	Oceanographic geographical features
OGC	Open Geospatial Consortium
OI	Orthoimagery
PA	Protected Areas
PD	Population distribution – demography
PF	Production and industrial facilities
POLIMI	Politecnico di Milano, Dipartimento di Elettronica e Bioingegneria
Provita/IUCN	Provita, International Union for Conservation of Nature
PS	Protected sites
QA	Quality Assurance
QC	Quality Control
QI	Quality Indicators
QualityML	Quality Indicators Dictionary and Markup Language
REDIAM	Agencia de Medio Ambiente y Agua

REV!GIS	Resolving uncertainty in Geographical Information Systems
RFC	Request for Comments
RS	Coordinate reference systems
RS	Remote Sensing
SD	Spatial data
SD	Species distribution
SDI	Spatial Data Infrastructure
SDQ	Spatial data quality
SDQE	Spatial Data Quality Evaluation
SEIS	Shared Environmental Information System
SIU	Szent István University
SO	Soil
SR	Sea regions
STARLAB	Starlab Barcelona
SU	Statistical units
Task 5.5	Development of data quality evaluation routines
TdV	Station Biologique de la Tour du Valat
TDWG	Taxonomic Databases Working Group
ThemisE	**THE**matic **M**etadata-based and f**I**tness-for-use **S**patial data quality **E**valuation
TIN	Triangulated Irregular Network
TN	Transport networks
UAB	Universitat Autònoma de Barcelona
UB	University of Bucharest
UBO	University of Western Brittany
UBT-Bayceer	University of Bayreuth, Bayreuth Center of Ecology and Environmental Research
UFZ	Helmholtz Centre for Environmental Research
UiB	Universitetet i Bergen
UKT2	Terradue UK Ltd.
UML	Unified Modeling Language
UNEP	United Nations Environment Programme

UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIGE	University of Geneve
UNILE	Università del Salento
UNIVLEEDS	University of Leeds
UNSW	University of New South Wales
UPotsdam	University of Potsdam
URI	Uniform/Unique Resource Identifier
URL	Uniform Resource Locator
US	Utility and governmental services
UTF-8	8-Bit Unicode Transformation Format
Uumea	Umeå Universitet
VGI	Volunteered Geographic Information
VRE	ECOPOTENTIAL Virtual Laboratory Platform
WebGIS	Web Geographic Information Systems
WISE	Water Information for Europe
WP	Work packages
WP1	WP1: “Coordination and management”
WP10	WP10: “ECOPOTENTIAL Virtual Laboratory Platform”
WP11	WP11: “EO supported policy development and integration”
WP12	WP12: “Capacity building and knowledge exchange”
WP2	WP2: “Conceptual Scientific Framework”
WP3	WP3: “Earth Observation Data and Processes Infrastructure”
WP4	WP4: “Earth Observation Data Generation and Harmonization”
WP5	WP5: “In-situ Monitoring Data”
WP6	WP6: “EO based Ecosystem Modelling”
WP7	WP7: “Ecosystem Services”
WP8	WP8: “Cross-scale interaction”
WP9	WP9: “Requirements of future protected areas”
XML	Extensible markup language

Summary

The increasing amount of spatial data, and of users and uses of spatial data, has been associated with the development of systems and spatial data infrastructures within the framework of big, open and linked data. This trend has been observed in connection to environmental/ecological monitoring programs as well as to global and European research initiatives focused on biodiversity and ecosystem services.

Changes in the life cycle of spatial data - namely in terms of data modelling, management and sharing - raise critical challenges for internal and external spatial data quality evaluation (ISO 19157) and management (ISO 19158). Therefore, we reviewed the development and adoption of concepts, spatial (meta)data standards and data quality evaluation approaches, methods and tools (section 2). A broad range of recent research (basic and applied) highlights significant advances in tools for indirect, partial and external spatial data quality evaluation, using different spatial metadata profiles and catalogues.

The implementation and improvement of spatial data evaluation processes involved a comparative analysis of candidate spatial metadata profiles, in order to propose extensions in metadata fields and user-oriented quality attributes, as well as of the associated metadata cataloguing and management (section 3). The definition of user-oriented and fitness-for-use data quality evaluation methodologies and routines justified the setup of a survey supported by an online questionnaire to inquire the ECOPOTENTIAL (ECOP) community about the spatial data quality user´s knowledge, interests, experiences and recognized utility. The results from the survey (also presented in section 3) showed the limitations of ECOP researchers' in the knowledge, experience and quality evaluation practices, in contrast to their manifest interest in the application of routines and interest to expedite methods of data quality evaluation and management.

The theoretical introduction and literature review (section 2), together with the evaluation of metadata profiles as well as the knowledge and interests of ECOPOTENTIAL researchers (section 3), led to the design of a new **THE**matic **M**etadata-based and f**I**tness-for-use **S**patial data quality **E**valuation (ThemisE) platform (Section 4). This platform aims to support the quality-driven discovery and selection of relevant data (or the identification of data gaps) necessary for environmental/ecological modelling based on well documented datasets. The ThemisE platform has been developed as an autonomous and modular Web application, and includes functionalities to perform internal and external (meta)data quality evaluation. This is aimed to support the discovery of datasets that meet the user’s requirements, for which the quality of datasets from configured catalogues is evaluated by determining the matching level (fitness-for-use) between the characteristics of the dataset (detailed by its metadata) and the characteristics of the data required by the user (defined through expected values for predefined quality indicators). Additionally, the metadata quality of each dataset is analysed regarding its compliance with required elements and implementation rules for predefined metadata standard profile(s).

The ECOPOTENTIAL project (WP5 and beyond) provides a suitable context to test the ThemisE platform supporting practical modelling exercises and knowledge network management in the context of a collaborative research community. The integration of the platform with other tools developed by WP5 (metadata catalogue) and WP10 (Virtual Laboratory) will foster data quality management in the broader ECOPOTENTIAL community and contribute to expand these principles to e-science initiatives.

1. Introduction

The ECOPOTENTIAL project combines Earth Observations from remote sensing and field measurements, data analysis and modelling to estimate current and future ecosystem conditions and services, with an emphasis on Protected Areas (PA). ECOPOTENTIAL considers cross-scale geosphere-biosphere interactions at regional to continental scales, while recognizing that anthropogenic pressure has caused serious threat to ecosystem integrity, functions and processes. Knowledge-based conservation, management and restoration policies are thus urgently needed. Effective monitoring and modelling of the state and trends in ecosystem conditions and services are crucial to this endeavour (http://www.ecopotential-project.eu). The project is structured in twelve thematic work packages (WP), each subdivided into specific Tasks (Fig. 1-1).

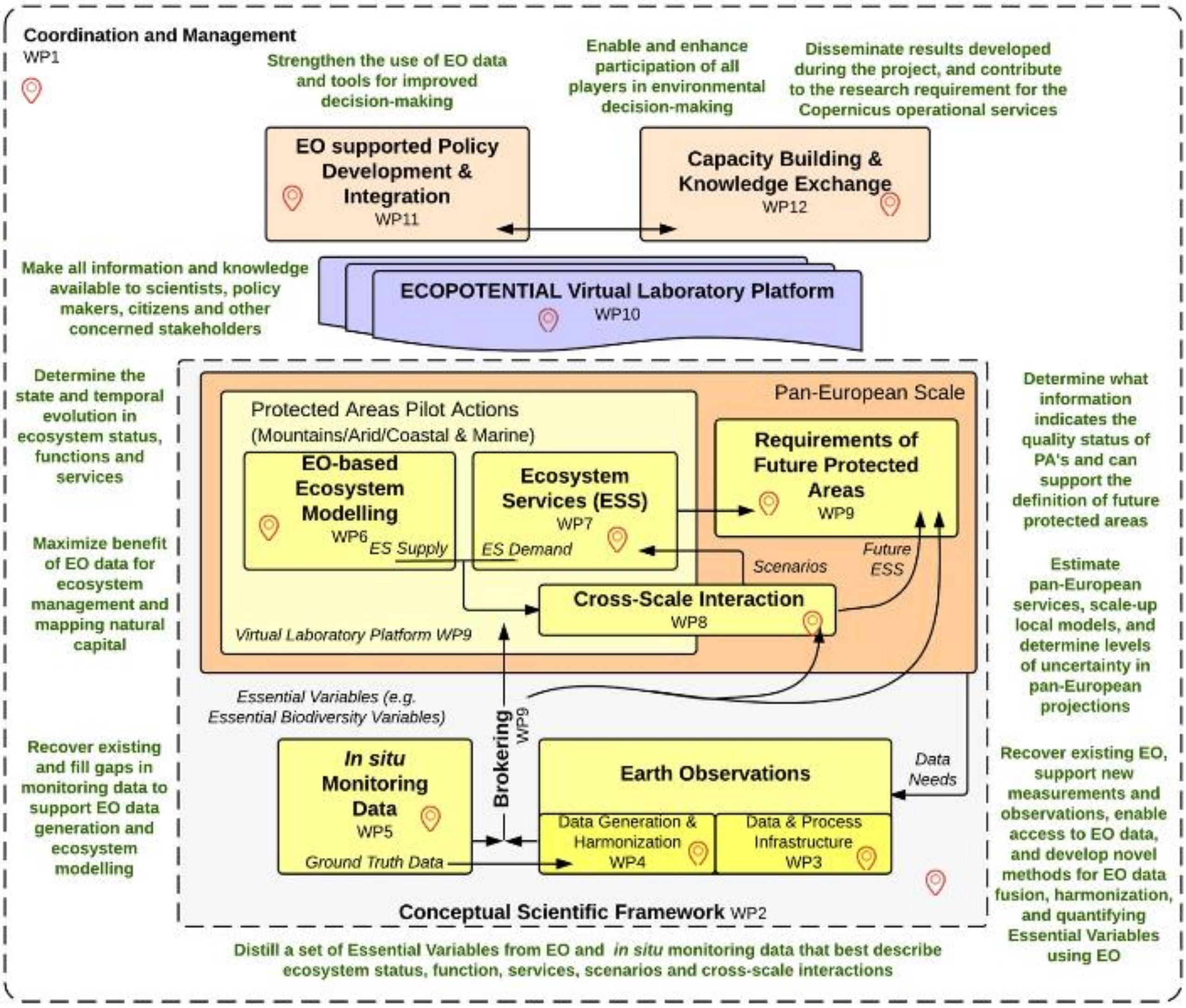


Fig. 1-1. ECOPOTENTIAL Project Framework (refer to the website or the project proposal).

A major effort of ECOPOTENTIAL is dedicated to make existing EO usable and inter-operational, to support new algorithms to recover EO data, and to create ecosystem-relevant knowledge. In parallel to the EO-focused effort, ECOPOTENTIAL emphasizes on the recovery, mobilization and gap-filling of in situ monitoring data. This is an essential prerequisite for achieving credible EO products, ecosystem models and estimates of ecosystem services. In this context, the aim of WP5 is to prepare and provide access to existing relevant in-situ monitoring data, in order to support data analysis workflows (e.g. model calibration, validation of remote sensing products). This includes the preparation of metadata as well as the physical provision of data. WP5 focuses its activities on the following objectives:

(i) Identify and review relevant core and ancillary in-situ datasets based on project requirements (e.g. modelling approaches, ground-truthing or validation of EO data) and identify gaps (Task 5.1);

(ii) Identify and provide datasets on the historic evolution of ecosystems based on proxy data as input to the assessment of the effects of historic conditions to current ecosystem changes (Task 5.2);

(iii) Provide complete, consistent and standard compliant metadata for data quality assessment, data management and data analysis (Task 5.3);

(iv) Evaluate the representativeness of existing data for PAs e.g. in terms of coverage, grain, or accessibility in order to identify knowledge gaps and priority areas for EO (Task 5.4);

(v) Develop user-oriented quality evaluation routines and procedures (data quality management) and web based quality evaluation software for in-situ data (Task 5.5);

(vi) Provide a framework for data harmonization (temporal, spatial, and thematic dimensions) for relevant data related to data management and interoperability for time-series and geospatial data (Task 5.6); and

(vii) Create a database for the integration of metadata from different sources (Task 5.7).

Task 5.5 is devoted to develop and implement user-oriented quality evaluation routines for an agile and adequate assessment of the internal and external quality of pre-existing (spatial) data based on their metadata (see Task 5.3), following the principles of ISO 19157:2013 and ISO 19158:2012. This task benefits from the experience and results from previous European projects (e.g. BIO_SOS, EU BON) (inter-alia developing and testing (Alonso et al., 2013a; Honrado et al., 2011; Pôças et al., 2014), fitness-for-use evaluation tools using open source software. Such tools support the quality-driven selection of relevant data identified as well as the identification of data quality gaps and the planning of targeted data collection.

Task 5.5 is led by ICETA/InBIO and involves interaction with several other ECOPOTENTIAL Tasks and partners (Fig. 1-2). The activities included in this task are:

(i) A literature review on methodologies and routines for user-oriented data quality evaluation. This first activity involved the search, collection and review of literature concerning spatial data quality (SDQ) assessment and management, including: (1) quality concepts, domains, references and challenges; (2) spatial data quality assessment; and (3) user-oriented (spatial) data quality assessment/evaluation methodologies and routines in the context of knowledge network management.

(ii) The identification, analysis and specification of requirements for the user-oriented quality evaluation routines, supported on metadata. Activities developed under this topic included: (1) analysis of several candidate metadata profiles (INSPIRE, DEIMS Community Profile, EML, ISO19115, ISO 19157, …); (2) evaluation of the current DEIMS-SDR DataSet MD model in terms of adequacy for data quality assessment; and (3) proposal of new fields for the ECOPOTENTIAL Community Metadata profile, related to data quality elements (in cooperation with Task 5.3). In the context of task 5.5, activities are focused on the user's input data quality requirements according to the application context (spatial modelling) and to the users' expertise (or "expected quality"). Therefore, an online questionnaire was developed (also in cooperation with task 5.3) around which quality elements would be more useful/important for data providers and for data users.

(iii) Specification of the requirements for data quality routines and extraction of metadata from the ECOPOTENTIAL VRE (linking to task 5.7), which includes the development of service-based routines in order to create a modular system that facilitates the integration with the global platform, as well as testing. These activities should include all partners involved in the task, as well as all partners interested in applying spatial data quality assessment and management (e.g. to their PA databases). This resulted in the development of a Web SDQ fitness-for-use platform, using open source software, in which fitness-for-use evaluation tools are developed and tested with a strong input from data users. A module for user-oriented quality evaluation was developed with multiple features. This module was developed using some of main web languages, including HTML, CSS and JavaScript, allowing to be used in any browser chosen by the user.

(iv) Cooperation with Task 5.7 and with WP10 around the integration with the database framework and with the service-based platform (the ECOPOTENTIAL Virtual Laboratory Platform), especially on how to get the metadata from the system, and how to create data quality routines.

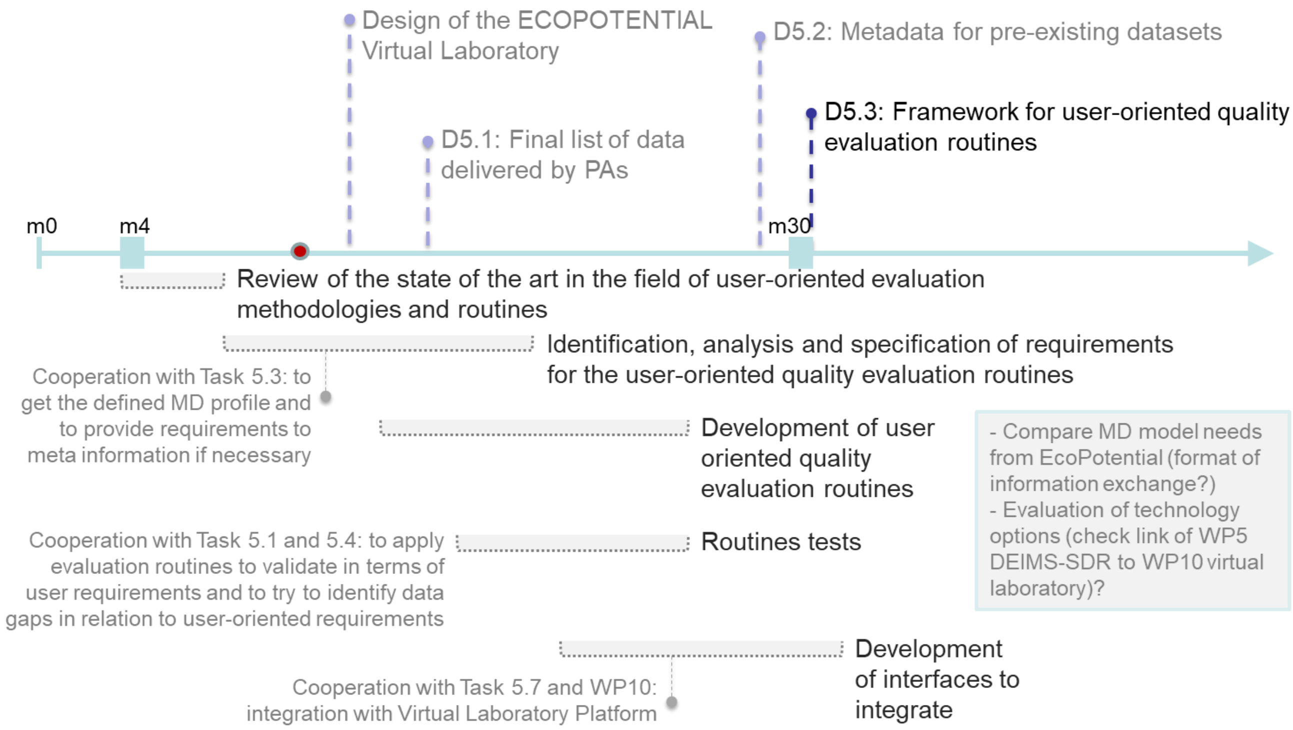


Fig. 1-2. WP 5.5 work plan on WP5 framework and other related (sub)tasks and ECOPOTENTIAL deliverables.

The Task W5.5 approach, activities and results are centred on in-situ data, including EO data (WP5). The spatial nature of these data allows the methodologies, routines and tools obtained and described in this Deliverable (D5.3) to be applied and extended to other images / EO data revisiting specific aspects of standards such as ISO19115: 2 (Geographic information - Metadata - Part 2: Extensions for imagery and gridded data) and ISO 19139-2: 2012 (defines Geographic Metadata for imagery and gridded data (gmi) encoding).

This report (**Deliverable D5.3**, “Framework for user-oriented quality evaluation routines”) describes the most important results of the activities developed in the context of Task 5.5, including: (i) A literature review on the state of the art of spatial quality assessment based on metadata (section 2); (ii) The proposal of new fields to expand the selected metadata profile, related to data quality elements and allowing more effective spatial data quality assessment based on metadata entries (collaboration between task 5.3 and task 5.4) (section 3.1); (iii) The results of a survey, supported by an online questionnaire on “knowledge and routines of data quality assessment and management”, to inquire the ECOPOTENTIAL community of data providers and data users on their experience and awareness of data quality routines (section 3.2); and (iv) The collaborative specification and development of the new Web SDQ fitness-for-use ThemisE platform using open source software (section 4).

2. Methodologies and routines for user-oriented data quality evaluation

Knowledge and conservation of biodiversity and ecosystems are critical elements for land habitability and sustainability. In recent decades, there has been an exponential increase in the generation, access and use of (spatial) biological, ecological and environmental data. This calls for effective advances in the evaluation and management of data quality, aimed at improving communication, decision support and knowledge management in open, linked and collaborative contexts. The current section focuses on the referential concepts and discusses approaches to developing methods and instruments for evaluation and management of the quality of spatial data to be tested in ECOPOTENTIAL.

This section frames the relevance of developing data quality evaluation procedures in ECOPOTENTIAL, namely considering that: (1) novel challenges in spatial data quality assessment are raised by the increasing amount of spatial data production (e.g. earth observations, in-situ measurements, model predictions and simulations); (2) spatial data are used in several different application contexts (i.e. ecosystem service assessments, conservation of natural heritage); (3) neglecting data quality fosters the risk of misuse, misinterpretation and can cause misleading results; and (4) data quality must be an essential criterion to identify datasets that satisfy the requirements of a particular application for a specific user (e.g. when developing statistical/correlative or process-based ecosystem models in WP6). Therefore, measuring, assessing, managing and communicating (spatial) data quality is important throughout the data life cycle and associated data processing, analysis, communication and decision-action processes.

2.1 Spatial data quality concepts, references and challenges

Spatial data contribute to multidisciplinary knowledge development, improve functional systems analysis promoting understanding of structures and processes (Guerra et al., 2010). Spatial data, models and spatially explicit systems also facilitate communication, participation and the development of technical-political decision support systems. Data digitization has promoted production, required new requirements in data encoding (data structures and models), storage, processing, access and data sharing, within a framework of increasing demands on data security and privacy, data/processes standards and system/infrastructures information advances.

The scientific knowledge, technological and organizational innovation and policy options frame the development and diffusion of Geographic Information Systems (GIS) to implement Spatial Data Infrastructure (SDI) at a global, national and local scale as well as, disciplinary and communities level related to global/local project, initiatives and knowledge networks. GIS and SDI integrate data, technologies, users, standards and policies in the production, management, sharing and application of spatial data and geographic information. The GIS focuses on the data production and analysis, while the SDI prioritizes communication, sharing and easy, adequate and secure access to data and data services between users and systems.

SDIs are digital information infrastructures that promote digital governance initiatives, spatially enabled societies and communities led by public institutions for participation and social cohesion and inclusion, environmental quality, land sustainability and the generation of new knowledge economies by private entities. Under intense climate and socio-ecological changes, the importance and challenges of habitability and land sustainability imply environmental observation and monitoring (GEOSS, COPERNICUS, Earth Systems Data Cube) as well as environmental/ecological modelling (EM) relating of biodiversity management and ecosystem services assessment (BISE, GBIF, LIFEWATCH, IPBES). These purposes require the increase of (spatial) data capture, modelling, management, sharing and access (Big, Open and Linked Data) included in thematic applications, systems and (cyber)geographic information infrastructures development framework.

The recognized advantages of WebGIS and SDIs have promoted integrative or thematic initiatives from global to local level (Rajabifard et al., 2010). Since 2003, with the creation of the intergovernmental Group on Earth Observations (EO) (http://www.earthobservations.org), and 2004, with a worldwide commitment for the implementation of the Global Earth Observation System of Systems (GEOSS) (http://www.earthobservations.org/geoss.shtml), governments have recognized the key role of Earth observation and the urgent need for a combined effort to identify, characterize and evaluate global change and its effects on components of human well-being[1]. In this sense, biodiversity should represent one of the main subsets of such an Earth observation infrastructure.

Also, at the European level several initiatives are developing, namely: (i) COPERNICUS[2] (http://www.copernicus.eu/); (ii) INSPIRE - Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007, establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)[3] (http://inspire.jrc.ec.europa.eu/); or (iii) Shared Environmental Information System (SEIS/GALILEO)[4] (http://ec.europa.eu/environment/seis/). In recent years, these initiatives have tried to promote and integrate global/European initiatives and thematic SDIs (e.g. Water Information for Europe - WISE), SDI in nature conservation (NatureSDIplus) (http://www.nature-sdi.eu/), the Biodiversity Information System for Europe (BISE) (http://www.eea.europa.eu/publications/bise), Environmental monitoring and biodiversity spatial databases and thematic SDIs like the Global Biodiversity Information Facility (GBIF) (https://www.gbif.org/), e-infrastructure for biodiversity and ecosystem research LIFEWATCH (http://www.lifewatch.eu/), Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (https://www.ipbes.net/); long-term ecological research network LTER NETWORK (https://lternet.edu/)). This includes also the promotion of the stakeholder's involvement, data sharing and disseminates best practices.

All these initiatives facilitate the implementation of spatially explicit ecological or environmental monitoring programs, which are crucial for the gathering and consolidation of knowledge related to the patterns of distribution, function, and interaction of biological assets with other spatially explicit factors (e.g. land cover, human development, and environmental disasters) related to ecosystems services modelling, ecological/environmental conservation and land management.

Initiatives or projects involving the handling of high volumes of geospatial data, with a similar or complementary scope to ECOPOTENTIAL, typically use methods for assessing, measuring, reporting and controlling spatial data quality (European Spatial Data Infrastructure with a Best Practice Network (ESDIN) -

[1] One of the main goals of GEOSS is to link existing systems and networks to achieve comprehensive, coordinated and sustained observations of the Earth system. In order to accomplish this, efforts must be put into implementing, standardizing and evaluating existing data flows and infrastructures to promote better communication between observation systems, in agreement with political, legal, organizational and standard references associated to Global Spatial Data Infrastructure (GSDI) development.

[2] Copernicus, previously known as GMES (Global Monitoring for Environment and Security), is the European Programme for the establishment of a European capacity for Earth Observation. This is the European Program for the establishment of European capacity for Earth Observation services, addressing six main thematic areas: Land Monitoring, Marine Environment Monitoring, Atmosphere Monitoring, Emergency Management, Security, and Climate Change.

[3] aims to ensure that spatial data infrastructures of the Member-states are compatible and usable in a Community and transnational context; the Directive requires that common Implementing Rules are adopted in a number of specific areas (Metadata, Data Specifications, Network Services, Data and Service Sharing, and Monitoring and Reporting.

[4] Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - Towards {SEC(2008) 111} {SEC(2008) 112}) - this Communication sets out an approach to modernize and simplify the collection, exchange and use of data and of information required for the design and implementation of environmental policy; the overall aim is to maintain and improve the quality and availability of information required for environmental policy, in line with better regulation, while keeping the associated administrative burdens to a minimum.

http://www.esdin.eu/). The increasing quantity, diversity and heterogeneity of (pre-existing and new) reference and thematic spatial data are associated with intense and critical data life cycle changes namely at recent (spatial) data-warehouses collection/production (e.g. static and mobile ground sensors/environmental facilities, aerial and space sensors/images….) and data processing/sharing advances (e.g. volunteered geographic information - VGI, citizen and collaborative science, Internet of things- IoT[5]) contributing to: (i) increased demand for data quality assessment, processes and knowledge management; (ii) reinforced training opportunities/education, and individual/organizations capacity building needs; as well as (iii) improved communication and technical-political decision processes and social participation/inclusion.

Interoperability and data sharing challenges, among a growing diversity of (multi)users, distributed, multitask and multifaceted environment imply to specify data models, concepts and standards (ISO 19100; OGC standards[6] and INSPIRE data models specification), develop tools (e.g. metadata catalogue, browsers/search engine) and implement procedures (e.g. spatial data quality assessment) that guarantee spatial data documentation, access and communication as well as, support spatial data quality management processes (control and assurance). Therefore, data quality is a very active domain in geographic information science/research in these last thirty years that´s accompanied and advance with geographic information science and technology development (Devillers et al., 2007; Goodchild, 2009; Longley et al., 1999; Goodchild, 1995).

Quality is defined by ISO 8402 as the "*totality of characteristics of a product that bear on its ability to satisfy stated or implied needs*" (ISO, 1994), and by ISO 9000 as the "*degree to which a set of inherent characteristics fulfils requirements*" (Brando and Bucher, 2010; Ingberg, 2006). Analogic or digital spatial/geographic datasets/database models describe the real world from different, complementary and alternative possible viewpoints/perspectives (data producer's/provider's universe of discourse related to features/elements selection/abstraction with the definition of symbols and spatial/thematic description), giving direct/absolute (spatial coordinates) or indirect/relative (name, administrative, statistical or postal area) reference spatial location. Human perceptions, knowledge, competences and limited aptitudes as well as technological and human data handling errors induces differences/uncertainty between complex reality and the simple/reduced data model, datasets and databases obtained (Liu et al., 2016; Longley et al., 1999).

The multiple and hierarchical reuse of chain spatial data offers increasing data availability and includes all users as part (responsible) of data life cycle. Spatial data are increasingly being transformed, shared and used for other purposes different from those for which they were originally produced. Spatial data life cycle imply attention to spatial data quality elements, evaluation/assessment procedures and quality indicators

[5] Associated to socio-ecological/environmental long term-monitoring initiatives.

[6] These advantages and opportunities have led to the conceptualization, proposal, discussion and application of different **global standards**:

(i) **ISO 19100** is a series of standards for defining, describing, and managing geographic information, *i.e.* information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth; this series of standards specifies methods, tools and services for management of information, including the definition, acquisition, analysis, access, presentation, and transfer of such data in digital/electronic formats between different users, systems and locations; these standards make it possible to define profiles in order to facilitate the development of geographic information systems and application systems that will be used for specific purposes; in this context, "*profiling*" consists of putting together "*packages*" or "*subsets*" of the total set of standards to fit individual application areas or users (ISO 19100 Series of Geographic Information Standards, 2004).

(ii) The **Open Geospatial Consortium** (OGC) is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services; OGC standards are technical documents that detail interoperability guidelines; software developers use these documents to build open interfaces and encodings into their products and services; these standards are the main "*products*" of OGC and have been developed by its members to address specific interoperability challenges; ideally, when OGC standards are implemented in products or online services by two different software engineers working independently, the resulting components will plug and play, that is, they will work together without further debugging (OGC, 2011). **Benefits** of using (and enforcing) data standards include: (i) more efficient data management (including updates and security); (ii) increased data sharing; (iii) higher quality data; (iv) improved data consistency; (v) increased data integration; (vi) better understanding of data, and (vii) improved documentation of information resources (National Land & Water Resources Audit, 2008).

(Devillers et al., 2007). The contexts within which geospatial data are used have changed significantly. Users now have easier access to geospatial data but often have less knowledge in geographical science and spatial technological domain. Most spatial data users have inadequate perception knowledge and competences of spatial data quality limitations that induces data handling, modelling and analysis. This risks the reduction of the relevance of the results and induces bias and uncertainty on communications and decisions (Devillers et al., 2007). van Oort (2006) refers to five main reasons for current concerns and challenges about spatial data quality issues: (i) there is an increasing production, availability, access, sharing and use of spatial data across expert data providers (national or national reference authorities) and (non)expert multidisciplinary data users; (ii) there is a growing number of users less aware/prepared of spatial data quality assessment and management; (iii) GIS enable/permit the spatial data use in several increasing applications, regardless of the appropriateness with regard to data quality aware; (iv) current GIS offer hardly tools for handling spatial quality; (v) there is a physical, disciplinary and professional increasing distance between those who use the spatial data (the final users) and those who are best informed about the quality of the spatial data (the data providers).

The ground, aerial and spatial (mobile) sensors (observations), as well as, spatially explicit data analysis and modelling (simulations) reinforce dynamic nature of vector and image spatial data time series' documentation and management needs. The development of technological and data sharing protocols refer to the relevance of data times series collection, data processing transformation and data sharing quality evaluation as well as, the critical importance of data (geo)web services quality assessment. The recent spatial data life cycles changes reveal quality assessment and management trends focusing on the thematic users and data advisors, reinforcing user-side thematic data and user quality perspectives as well, highlight the challenge of assessment and ensure quality throughout processes. Spatial data quality assessment and management imply conceptual and methodological advances and technological challenges, human capacity building, political options and social organization considering the tetrahedron of quality, between reality and their analogue/digital representation, between producer and users (Fig. 2-1).

Fig. 2-1. Spatial data quality scope, analysis and management at GIS/SDI development framework (Jakobsson, 2011).

Co-funded by the European Union

2.2 Spatial data quality assessment processes and tools

Data quality is relevant to describe, understand and support decision about reality, as well as to promote efficient communication, inclusion and individual or collective human decision/responsibility (Devillers and Jeansoulin, 2006). Spatial data quality assumes different perspectives along spatial data life cycle from the data models requisites analysis and product specification phase (dedicated to data product specification), chain and operation production phase (quality control and assurance), quality assessment (direct/indirect, total/partial and quantitative/qualitative assessment), and spatial dataset and database documentation/communication to potential users using metadata as data synthesis content and resulted quality evaluation reports.

The scientific and technological community of (spatial) data quality evaluation/assessment presents new conceptual approaches and methodological frameworks, methods and instruments/tools in the total or partial (in)direct spatial data evaluation namely, in the quality elements/indicators generation and representation of the spatial variability of quality elements. In this context, the data quality and data quality evaluation has received attention in geographical information science community with standards and guidance's implementation (Beare, 2010), as the recent International Standard ISO 19157:2013 – *Data quality* (DQ) (ISO, 2013), replaces ISO/TS 19138:2006 – *Data quality measures* (ISO, 2006), ISO 19114:2003 – *Quality evaluation procedures* (ISO, 2003), and ISO 19113:2002 – *Quality principles, and establishes the principles for describing the quality of geographic data* (ISO, 2002).

ISO 19157:2013 fulfils the need giving conceptual framework for the issue of data quality and for process data quality evaluating, and can be described "data quality" using different components: data quality elements and their descriptors (sub-elements) (Tab. 2-1); data quality measure (the type of evaluation); data quality evaluation (the procedure used to evaluated the measure); data quality result (the output of the evaluation); meta quality (the quality of the data quality results in terms of defined characteristics) (Tab. 2-1). Data quality elements (Tab. 2-1) include positional, thematic and temporal accuracy, completeness, logical and consistency. However, in last ISO 19157:2014 version, in addition to data quality elements and their descriptors, includes the term "Usability" (Tab. 2-1) which is described as: '*the degree of adherence to a specific set of quality requirements*', '*usability shall be used to describe specific quality information about a dataset's adherence to a particular application or requirements*' and '*the element may be used to declare the conformance of the dataset at a particular specification*', for example for a particular usage within a specific application (Leibovici et al., 2013).

ISO 19157:2013 also defines a set of data quality measures for use in evaluating and reporting data quality. It is applicable to data producers providing quality information to describe and assess how well a data set conforms to its product specification and to data users attempting to determine whether specific geographic data are of sufficient quality for their particular application. ISO 19157:2013 integrates and simplify preceding standards, and establishes (i) the concepts for handling quality information for geographic data; (ii) the components and structures of data quality measures; (iii) defines commonly used data quality measures; and (iv) provides guidelines on how to describe, evaluate and report data quality (ISO, 2013). This standard does not attempt to define minimum acceptable levels of quality for geographic data.

Tab. 2-1. Data quality element (ISO, 2013).

DQ_Element/*Sub-element*	**Description**
DQ_Completeness	presence and absence of features, their attributes and relationships
DQ_CompletenessCommission	*excess data present in a dataset*
DQ_CompletenessOmission	*data absent from a dataset*
DQ_LogicalConsistency	degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical). If these logical rules are documented elsewhere (for example in a data product specification) then the source should be referenced (for example in the data quality evaluation)
DQ_ConceptualConsistency	*adherence to rules of the conceptual schema*
DQ_DomainConsistency	*adherence of values to the value domains*
DQ_FormatConsistency	*degree to which data is stored in accordance with the physical structure of the dataset*
DQ_TopologicalConsistency	*correctness of the explicitly encoded topological characteristics of a dataset*
DQ_PositionalAccuracy	accuracy of the position of features within a spatial reference system
DQ_AbsoluteExternalPositionalAccuracy	*closeness of reported coordinate values to values accepted as or being true*
DQ_RelativeInternalPositionalAccuracy	*closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true*
DQ_GriddedDataPositionalAccuracy	*closeness of gridded data spatial position values to values accepted as or being true*
DQ_ThematicAccuracy	accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships
DQ_ThematicClassificationCorrectness	*comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data)*
DQ_NonQuantitativeAttributeCorrectness	*measure of whether a non-quantitative attribute is correct or incorrect*
DQ_QuantitativeAttributeAccuracy	*closeness of the value of a quantitative attribute to a value accepted as or known to be true*
DQ_TemporalQuality	quality of the temporal attributes and temporal relationships of features
DQ_AccuracyOfATimeMeasurement	*closeness of reported time measurements to values accepted as or known to be true*
DQ_TemporalConsistency	*correctness of the order of events*
DQ_TemporalValidity	*validity of data with respect to time*
DQ_UsabilityElement	usability is based on user requirements. All quality elements may be used to evaluate usability. Usability evaluation may be based on specific user requirements that cannot be described using the quality elements described above. In this case, the usability element shall be used to describe specific quality information about a dataset's suitability for a particular application or conformance to a set of requirements.

According to ISO19157: 2013, the process for evaluating data quality is a sequence of steps[7] to produce a data quality result, and used in different phases of a product life cycle, having different objectives in each phase. A data quality evaluation procedure comprises one or more data quality evaluation methods. Direct evaluation methods determine data quality through the comparison (full inspection[8] or sampling[9]) of the data with internal and/or external reference information. Indirect evaluation methods infer or estimate data quality using information on the data such as lineage (Tab. 2-2). The "Metaquality" describes the quality of the data quality results in terms of defined characteristics, such as "Confidence" - *trustworthiness of a data quality result*, "Representativity" - *degree to which the sample used has produced a result which is representative of the data within the data quality scope*, and "Homogeneity" - *expected or tested uniformity*

[7] Process steps for evaluating data quality: (i) specify data quality unit(s); (ii) specify data quality measures; (iii) specify data quality evaluation procedures; and (iv) determine the output of the data quality evaluation (ISO, 2013).
[8] Full inspection requires testing every item in the population specified by the data quality scope (ISO, 2013).
[9] Sampling requires testing sufficient items in the population in order to achieve a data quality result (ISO, 1989, 2013); sampling means that tests are performed on subsets of the geographic data defined by the data quality scope (ISO, 2013).

of the results obtained for a data quality evaluation (Fig. 2-2).

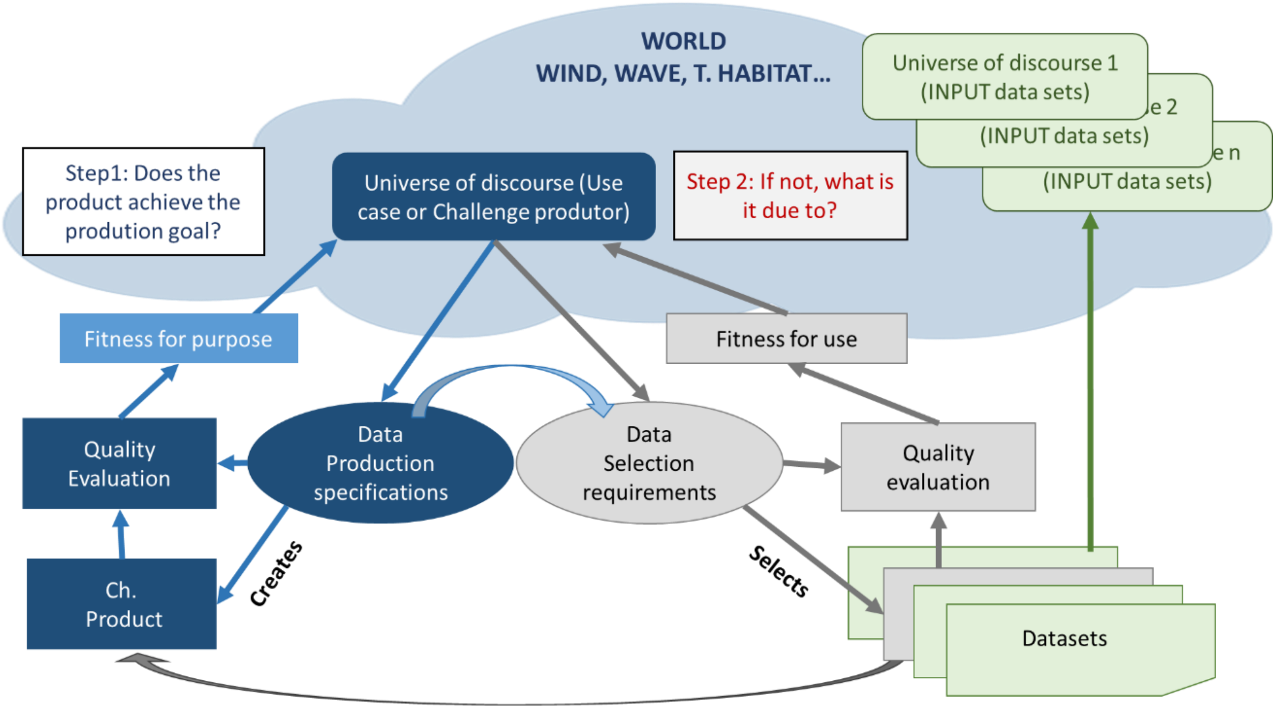


Fig. 2-2. Product specification and data user to dataset (ISO 19113 standard).

A quality evaluation process consists on the application of quality evaluation procedures to specific data, dataset or database related operations performed by the dataset producer and the dataset user. Processes for evaluating data quality are applicable to reference static and dynamic datasets (i.e., spatial data time series production). Dynamic datasets (including dataset times series) are datasets that receive updates so frequently that for all practical purposes they are continuously changing (EC, 2010; ISO, 2013). Data sharing implies evaluate the quality metadata (metaquality) and the quality of data (geo)web services between user and systems.

A possible workflow for evaluating and reporting data quality based on relevant ISO standards includes six steps recognized in a quality evaluation process (Fig. 2-3): (i) identify an applicable data quality element, data quality sub-element, and data quality scope[10]; (ii) identify a data quality measure, a data quality value type and, if applicable, a data quality value unit for each test to be performed; (iii) select and apply a data quality evaluation method for each identified data quality measure; (iv) determine the data quality result, i.e. a quantitative data quality result, a data quality value or set of data quality values, a data quality value unit and a date, as the output of applying the method; (v) determine conformance, whenever a conformance quality level has been specified in the product specification or user requirements; a conformance data quality result (pass-fail) is the comparison of the quantitative data quality result with a conformance quality level; (vi) report, i.e. the quality evaluation information shall be reported as metadata; a separate quality evaluation report is required when metadata result is only "pass/fail" or when aggregate quality results are generated (Fig. 2-3).

[10] This is repeated for as many different tests as required by the product specification or user requirements.

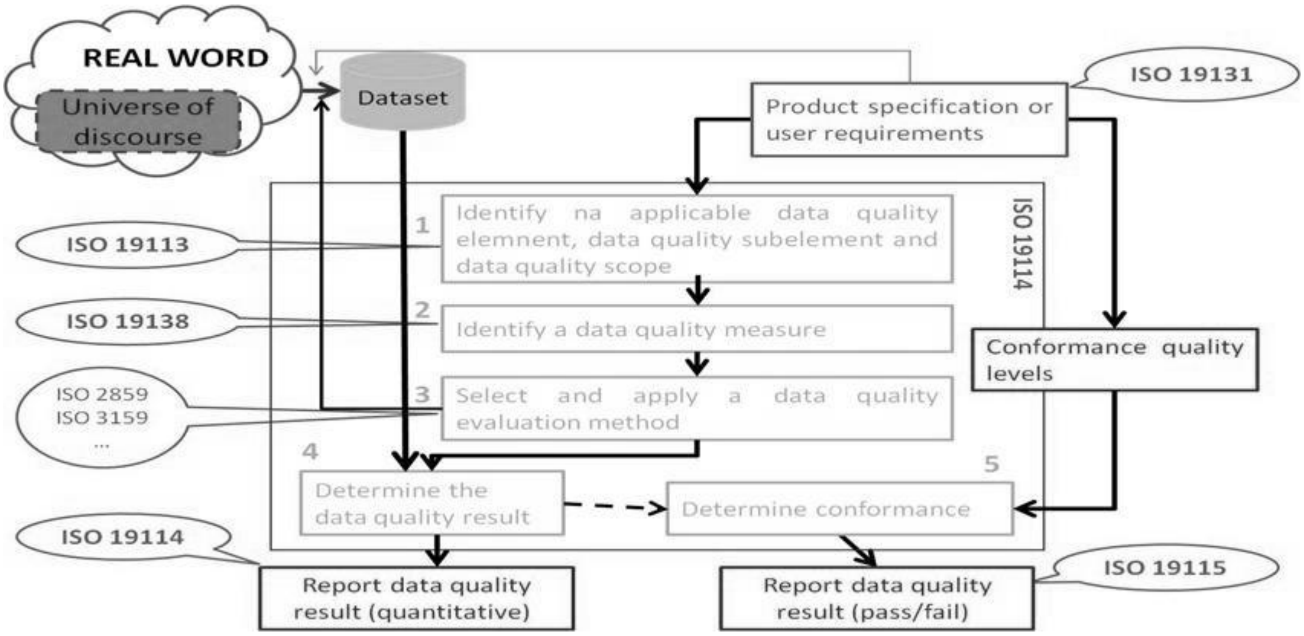


Fig. 2-3. Workflow for evaluating and reporting data quality results (Adapted from ISO/DIS 19114).

Direct evaluation methods are further subdivided into **internal** and **external**. All the information needed to perform an internal direct data quality evaluation method is inherent to the dataset being evaluated, while for external quality evaluation user defined requirements are needed. Therefore, the definition of data quality permits and gains with dual perspective, from product specification/development approach to "fitness for purpose" and on the user´s needs, distinguish quality in two broad and complementary internal and external assessment quality approaches (Devillers, Bédard, and Jeansoulin, 2005; Jakobsson and Tsoulos, 2007; van Oort, 2006; Vasseur, Devillers, and Jeansoulin, 2003) according to application context, user's interests and expertise "fitness for use".

Internal quality relates with the intrinsic characteristics of the data as described at the data provider level and usually detailed in metadata (Brassel, et al., 1995). This first level of assessment ("internal quality evaluation") is performed by the data producer/provider through a data quality check based on given preview data standard and technical guidance specifications. According to Devillers and Jeansoulin (2006), internal quality connects the quality of the data to the internal characteristics of the data, i.e., represents the difference between the produced data and intended and previous universe of discourse or defined model data. Internal quality presents a normative/standard approach.

Internal quality requires knowledge about procedures of spatial data quality management along spatial data life cycle, namely about: (i) data product specification; (e.g. ISO19131:2007 - Data product specifications, geographic data (https://www.iso.org/standard/36760.html)); (ii) data product specification user requirements and provide quantitative quality information (e.g. data models on Environmental Monitoring Facilities, Species distribution, Habitats and biotopes... (http://inspire.ec.europa.eu/data-specifications/2892)); (iii) collecting data (e.g. protocols; ISO 2854:1976 - Statistical interpretation of data (https://www.iso.org/standard/7854.html); ISO 3534-1:2006 - Statistics Part 1: General statistical terms and terms used in probability (https://www.iso.org/standard/40145.html); ISO 3534-4:2014 - Statistics Part 4: Survey sampling (https://www.iso.org/standard/56154.html)) about ancillary, historic and in-situ data which includes instrumental/technological component (e.g. instrumental selection and calibration) (relate to field surveys and environmental sensors facilities) as well, as thematic data (see INSPIRE Annex I, II and III) namely Remote Sensing/Earth Observation (RS/EO) data; (iv) data conceptual modelling; spatial analysis and modelling (e.g. data quality evaluation on Correlative and Process-based models); (v) data representation, publishing and sharing (e.g. analogic and digital graphical data communication relating data dissemination/accessibility) (e.g. ISO19109: 2015 - Rules for application schema, geographic data (https://www.iso.org/standard/59193.html)); and finally, (vi) evaluating data quality/reporting data quality (process data quality evaluating/metadata) (e.g. ISO19157:2013 - Data quality (DQ) (https://www.iso.org/standard/32575.html); QualityML - Quality Indicators Dictionary and Markup Language (Ninyerola et al., 2014).

External quality, generally recognized as the definition of quality in the broadest utility and usability sense, refers to the level of similarity between the data characteristics and the user's needs in a specific context of application, that is user-oriented or fitness for use quality (Devillers and Jeansoulin, 2006; Gervais et al., 2009). In synthesis, data quality assessment refers to reference data quality evaluation (i.e. normally direct and internal quality assured by data production chains assured by reduced number of cadastral/mapping authorities data providers and expert users) and external evaluation of thematic/disciplinary databases (normally realized by increasing number of multidisciplinary generic and non-expert/users). For both external and internal evaluation methods, there are two important choices to consider, automated vs. non-automated, and full inspection vs. sampling. A full inspection requires testing every item in the population specified by the data quality scope, whereas sampling requires testing sufficient items in the population in order to achieve a meaningful data quality result (Tab. 2-2).

Tab. 2-2. Classification of data quality evaluation methods (ISO, 2013).

Evaluation method type			**Means of accomplishing evaluation**
Direct evaluation method	Internal	All the data needed to perform an internal direct data quality evaluation method are internal to the dataset being evaluated.	- Full inspection - Sampling …
	External	External direct quality evaluation requires reference data external to the dataset being tested.	
Indirect evaluation methods	The indirect evaluation method is a method of evaluating the quality of a dataset based on external knowledge or experience of the data product and can be subjective. This external knowledge may include, but is not limited to one or more non-quantitative quality information usage, lineage and purpose (metadata), or other data quality reports on the dataset or data used to produce the dataset.		Data quality may be estimated: - Lineage - Metadata … (Metaquality)
Aggregation and derivation	Additional results may be produced by aggregating or deriving existing results without carrying out a new data quality evaluation. Aggregation combines quality results from data quality evaluations based on different data quality elements or different data quality scopes.		(Standalone quality report)

Therefore, assessing **quality** should include: (i) searching for a spatial dataset that contains the information needed for the intended application (Brassel et al., 1995 called this the "assessment of model completeness") (van Oort, 2006); (ii) exploring whether there are legal or financial constraints to access or to use the data

(Aronoff, 1989 called this the “usage component”) and linked/partial overlapped to usability (ISO 19115); and (iii) finding out if, given the quality of the data, risks are acceptable, adequate or critical to obtain or perform intended analysis (Agumya & Hunter, 2002).

External indirect quality evaluation can be implemented or improved by using metadata and metaquality concepts. Metadata, defined as ‘data about data’ or ‘information about information’, provides a fundamental basis for information management tools at three levels: (i) discovery, enabling users to locate and evaluate information; (ii) management, enabling custodians to better manage their spatial information; and (iii) utilization, enabling users to access and manipulate information by means of automated/distributed systems (Victorian Spatial Council, 2009). Metadata should include information on data quality as well as on the organizations responsible for providing the (meta)data management (ISO, 2013; Leibovici et al., 2013). In the first, based on user’s requirements, specific targets on data quality are established and have to be achieved in the course of data production or transformation. The second aspect corresponds to documenting the quality of the data that is eventually delivered to the users. For each of these tasks a common way of expression is necessary that comprises an agreed terminology, evaluation, and reporting methods (EC, 2010), both aspects leads to formalizing data quality. It´s critical to define adequate metadata to different reference and thematic data (ancillary, in situ, EO data, vector/raster and TIN data) relating to data models concepts/formats, production/processing and limited users perception.

It is therefore imperative that efficient and well-conceived, flexible and active metadata (or metaquality) standards exist and take into account data quality in the appropriate measure according dataset/databases scope/domain (Jakobsson and Giversen, 2008; Tóth et al., 2010) i.e., the definition of the metadata profile must incorporate quality quantitative and qualitative issues/fields of quality, useful both for internal and external quality evaluation. The purpose of describing the quality of geographic data facilitates the comparison and selecting of the dataset best suited to application needs or requirements. Information on the quality of geographic data allows a data producer to evaluate how well a dataset meets the criteria set forth in its product specification and assists data users in evaluating a product’s ability to satisfy the requirements for their particular application (Alonso et al., 2013b).

In recent years, efforts have been made in the meta-evaluation of external and (in)direct quality by the end-user(s), taking advantage of possibilities of documentation and communication of quality elements in metadata profiles. This requires simple and (semi)automatic metadata fulfilment, integration and metadata catalogue interoperability, in parallel, to the operationalization of the theoretical methods foreseen in the spatial data quality assessment and management standards (ISO 19115, ISO 19139, ISO 19157 and ISO 19158). Therefore, evaluation/assessment of external data quality of spatial datasets can be based on metadata (or data quality meta-evaluation): (i) metadata contains information about the content, quality, condition and other characteristics of data that can be used for external (meta)quality evaluation for knowledge discovery, indexing and searching; (ii) frequent limitations to data access and use advocate for SDQ evaluation based on metadata; (iii) availability and access of metadata catalogues allowing an (simple) integration of an evaluation methodology in (WebGIS) metadata catalogue platforms.

Metadata permit to describe, discover and communicate elements about (spatial) datasets and data quality elements in groups, communities and network contexts. Metadata enable the development and management of networks in scientific, technical and political open collaborative environments. Metadata improve the dialogue between the data user and the data producer by using elements that support the assessment of the spatial data quality relative to their intended application. Different metadata profiles (e.g. INSPIRE, DEIMS-SDR Dataset Community Profile, or DEIMS-SDR MD adopted in ECOPOTENTIAL, see section

3.1) provide limited information on data quality and need extension to cover this information. The EU REV!GIS project (Resolving uncertainty in Geographical Information Systems) suggest the use of metadata elements describing the operational use of data. These elements include information on (i) the socio-political context of data creation, the relevant actors and their influence; (ii) critiques on the data use (e.g. such as academic papers); (iii) the data producer´s opinions on the class separability (EO/RS data) and thematic accuracy; (iv) the expert opinions of relations to other datasets; (v) experiential metadata; and (vi) free text descriptions from producers; and tools for mining free text metadata (Pons & Masó, 2016).

At the same time it is critical to improve the management of metadata catalogues (e.g. sharing, maintenance, access and analysis), to promote metadata profile harmonization, to install and perform metadata services as well as, to perform (semi)automatic metadata fulfilment and enrichment. Current research challenges and technical advances are oriented to (semi)automatic fulfilment metadata as well as, to the publication and sharing of metadata at real/opportune time and in an agile manner. The importance of metadata is related to the information detail that is captured, while their production is time consuming and needs specific conceptual knowledge and technological competences for the researchers / data providers. Data provider's recognition and data user structural framework permits initial fulfilment of metadata elements and highlights the advantages of automatic edition as well as the regular and opportune updating of relevant data quality elements. Produce metadata and expanding metadata profiles/fields, namely data quality attributes, aims to close the gap between data providers and users as well as, produce and publish user and usability centred metadata. The metadata scope and intends should define metadata contents (related to data quality limitations), metadata formats and metadata (adequate) profile supporting standard metadata catalogues and services.

Recently, important progress was made to support the identification, analysis and specification of requirements for the user-oriented or fitness for use quality evaluation routines, supported on metadata (Bobillo et al., 2015; Ivánová et al., 2013; Pôças et al., 2014). To evaluate external data quality of spatial datasets based on metadata, it is necessary that metadata entries contain information about the content, quality parameters, access and use conditions, and other characteristics of datasets that can be used for external (meta)quality evaluation and also for knowledge discovery, indexing and searching. As such, efforts and advances have been made to ensure that the adequate metadata profile used, namely at quality oriented fields, will allow the application of user-oriented quality evaluation routines (Díaz et al., 2012; Goodchild, 2007).

2.3 User-oriented (spatial) data quality evaluation routines and knowledge network management

Data providers consider and are interested in the widest reuse of their data. This is only possible when the datasets satisfy the requirements of the targeted users, and the degree to which these requirements are fulfilled is documented. In the first, based on user's requirements, specific targets on data quality are established, and have to be achieved in the course of data production or transformation. The second aspect corresponds to documenting the quality of the data that is delivered to the users. For each of these tasks a common way of expression is necessary that comprises an agreed terminology, evaluation, and reporting methods. Both aspects lead to formalizing data quality.

External data quality, user-oriented an fitness for use data quality assessment methods and tools includes: (i) representation and inspection visual analysis relating to spatial error and uncertainty analysis inside dataset/database and/or comparative quality analysis between different datasets namely along data time series (Cuca et al., 2011); (ii) sometimes imply a direct access, assessment and inspection to original dataset/database limited by technological, property and private/individual rights; (iii) and increasing indirect, partial and external from user perspective data quality assessment exploring metadata catalogues and quality attributes/domains associated to select core, relevant and critical quantitative and qualitative methods and regular inspection moments (data control) under organizational, information system and database management framework.

Evaluation of external data quality of spatial datasets based on metadata refers to: (i) approach centred on the evaluation of similarities between the characteristics of the data according to users' requirements (or expectations) and the characteristics of the data as detailed by metadata (often described at producer level); (ii) users' requirements (perceived, recognized or expected quality) which consist of user-defined values for a set of specific quality indicators, according to application context and user's expertise; and (iii) the selection of quality indicators, which is based on the criterion that internal quality values could be filled in directly from information detailed in a metadata catalogue. Important progress was made in the activities to support the identification, analysis and specification of requirements for the user-oriented quality evaluation routines (Devillers et al., 2010).

In previous projects (e.g. BIO_SOS and IND_CHANGE) (Pôças et al., 2014; Alonso et al., 2013b; Honrado et al., 2011a) spatial data quality approaches and frameworks were developed (Fig. 2-4), which included metadata profile development, procedures and fitness for use and users requirements WEBGIS tools integrated with metadata catalogue management (Castro et al., 2013; Honrado et al., 2011b). A methodology was developed for evaluating external data quality supported by metadata entries, that are used to calculate the similarity between data characteristics (as documented by internal quality indicators stored in metadata fields) and user's needs or expectations (referred to as "expected quality" and highly dependent on the context of data usage) based on a process of geo-semantic integration. The method allows users to specify which quality indicators have stronger importance given their own requirements or expectations. The methodology is therefore based on the comparison between internal and expected values for selected quality indicators, using on a pre-defined rule-based system which controls how the pairwise comparisons are carried out (Honrado et al., 2011b; Castro et al., 2013; Pôças et al., 2014).

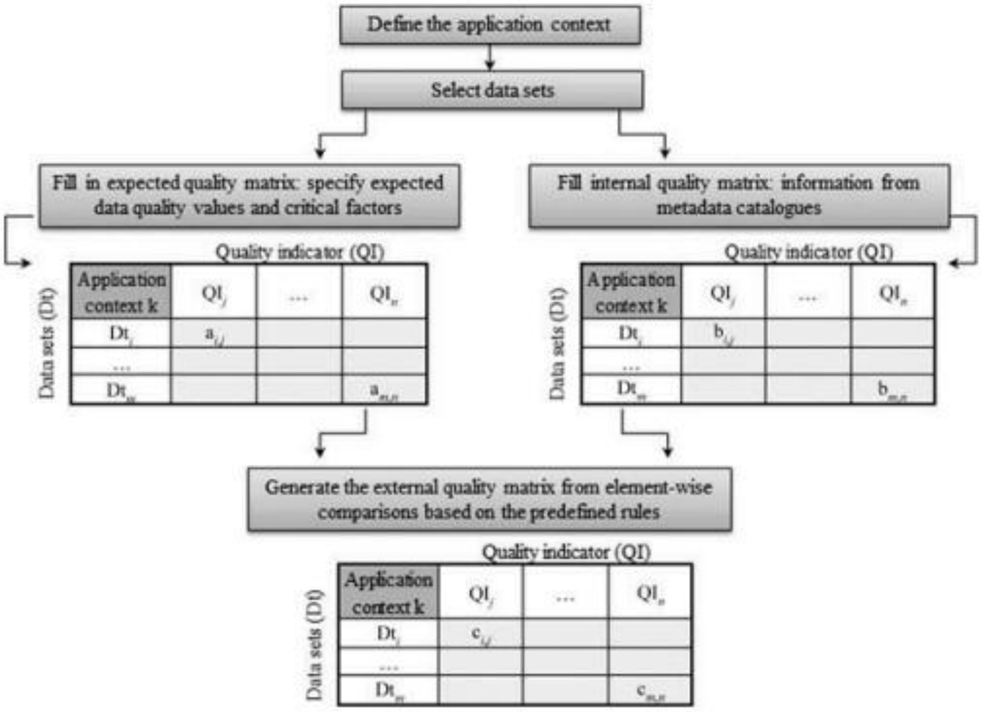


Fig. 2-4. General workflow of the external quality evaluation framework (EQDaM – BIOSOS Project) (Pôças et al., 2014).

Previous projects, research and tools considers different user-oriented (spatial) and fitness for use data quality assessment/evaluation methodologies and routines in the context of knowledge network management. Different related experiences select concepts and standard references (ISO 19157 and ISO 19158), internal and user-focused quality metadata, data quality methods and assessment tools along sequential phases/processes (Fig. 2-5), namely:

(i) the definition of an adequate metadata profile standards balanced between detailed (complete profile) and simple, relevant and critical fields/elements/attributes (core attribute metadata that are easily fulfilled) in order to perform spatial data quality evaluation (Lush et al., 2014);

(ii) the metadata filling and metadata management promoting easy and (semi)automatic completion of metadata (Manso-Callejo et al., 2008; Olfat et al., 2012) included in dynamic and active documented data life cycle management associated to organizational, information systems and infrastructures under knowledge network governance models and practices frameworks;

(iii) the load, organization, transformation, management, sharing and accessing to metadata catalogues linked to organizational/institutional dimension, SDI global, local and thematic initiatives constraints and information technological challenges (Díaz et al., 2012);

(iv) the definition, register and manage several (non)experts users requirements data resource profiles assuming different scientific background, geographic information science knowledge and experience (Curdt et al., 2014), interests and particular evolution needs/preferences (user-oriented) and relating specific uses defining intended quality (fitness for use) (Honrado et al., 2011a);

(v) the access, analysis, and cross-reference used methods between multifaceted user-specific profiles requirements and metadata quality oriented attributes related to find the datasets/databases available, adequate or optimum data quality needs during the search for spatial data resources (from datawarehouses, geoportals, or metadata catalogues);

(vi) the location and list exploring metadata fields, focused on core or critical metadata attributes (e.g. the platform identifies the spatial extent; application domain; …), hierarchize and compares alternative or complementary datasets/databases considering (fuzzy) multi-criteria analysis (qualitative and quantitative fields) and multi-objective analysis (users and use) from the users' needs/utility perception and quality risk analysis/management context (Sebake and Coetzee, 2013); the available and adequate data are complemented with the data gaps identification before access/download selected data (Honrado et al., 2011b);

(vii) exploring the representation (or visual inspection) of data quality results (description, numerical, graphical and spatial) according previous reports of data quality evaluation/Reporting data quality (process data quality evaluating/ metadata) (e.g. ISO19157:2013; QualityML);

(viii) the visualization, download, transform and manage relevant and suitable data for the intended use (fitness for use) according GIS/SDI data specifications standards (Waters et al., 2011), and/or intended software package requirements which implies (in)direct spatial, structure and data model transformation (e.g. INSPIRE data models specification);

(ix) the generation of new spatial data in spatially explicit modelling processes before fulfil new associated resulted metadata database and load to metadata catalogue assuming metadata enrichment and (meta)quality internal assessment from each standard metadata profile (Olfat, 2013).

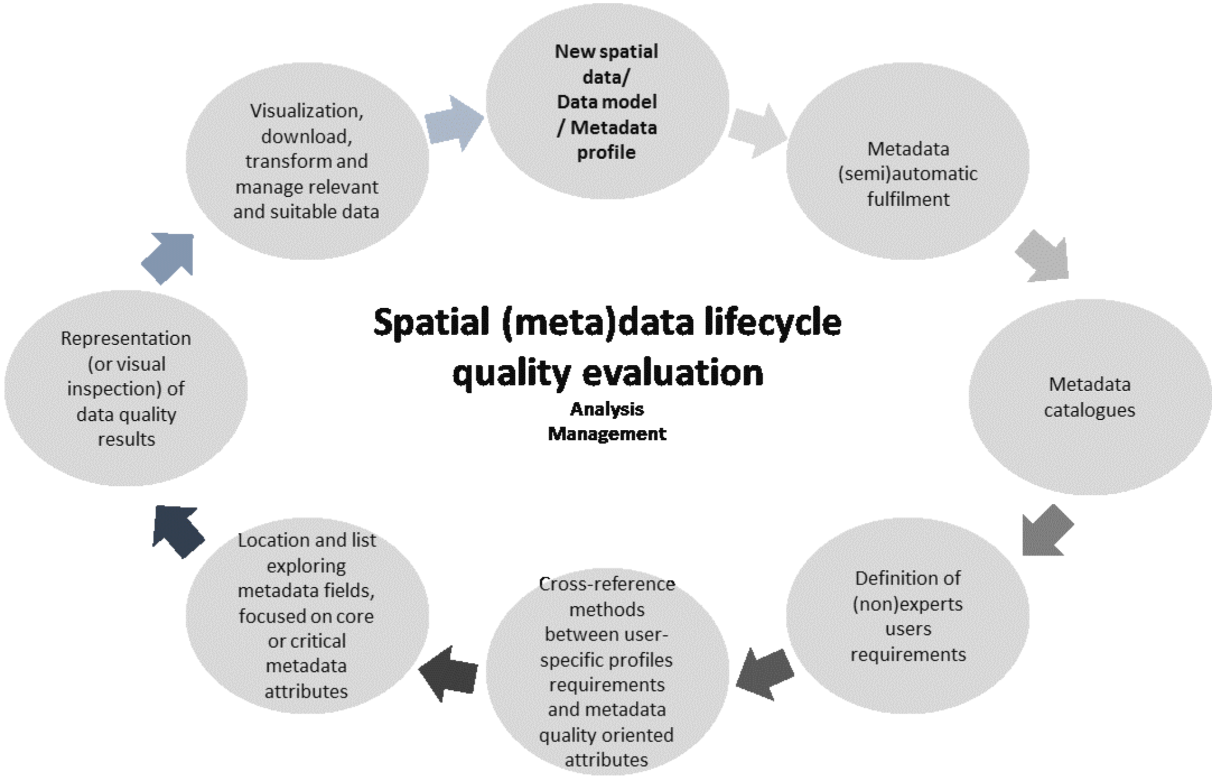


Fig. 2-5. Linked (meta)data life cycle and data quality assessment and management.

The discovery and access of spatial data resources imply the development of guided (web)search interfaces which explore and enhance search engines and consider logic decision on fitness for use crossing spatial (meta)data resources/catalogues and several multifaceted user's profiles (Fig. 2-6). In the indirect evaluation discussion should consider the development of "geospatial user feedback - GUF" in GeoViQua project (QUAlity aware VIsualization for the Global Earth Observation System of systems Project) transformed in OGC standard called GUF. In NextGEOSS (the next generation of the GEOSS), UAB (in collaboration with CREAF) is developing a user feedback system that allows to report about usage, data quality, comments, publications, ... between other relevant quality elements about the data. At same time, COBWEB (Citizen Observatory Web) was a research project that succeeded in its goal of developing a generic crowdsourcing infrastructure platform and toolkit that could be used in multiple scenarios generating data of sufficient quality to be used by policymakers (http://cordis.europe.eu/result/rcn/201513_fr.html).

This parallel and convergent intends refers to the need to develop simple and free-form search request and metadata cross-analysis and searches data resources that inform/comply/best fit quality requirements/user's needs. The development of spatially explicit models implies to choose and integrate adequate data and present results/effective proposals supported on data quality assessment and management principles and procedures. Data quality and data quality evaluation has received attention in geographical information science community (ISO/TC 211 Geographic information/Geomatics) with standards and guidance's implementation (http://www.geolabel.info/) (Beare et al., 2010).

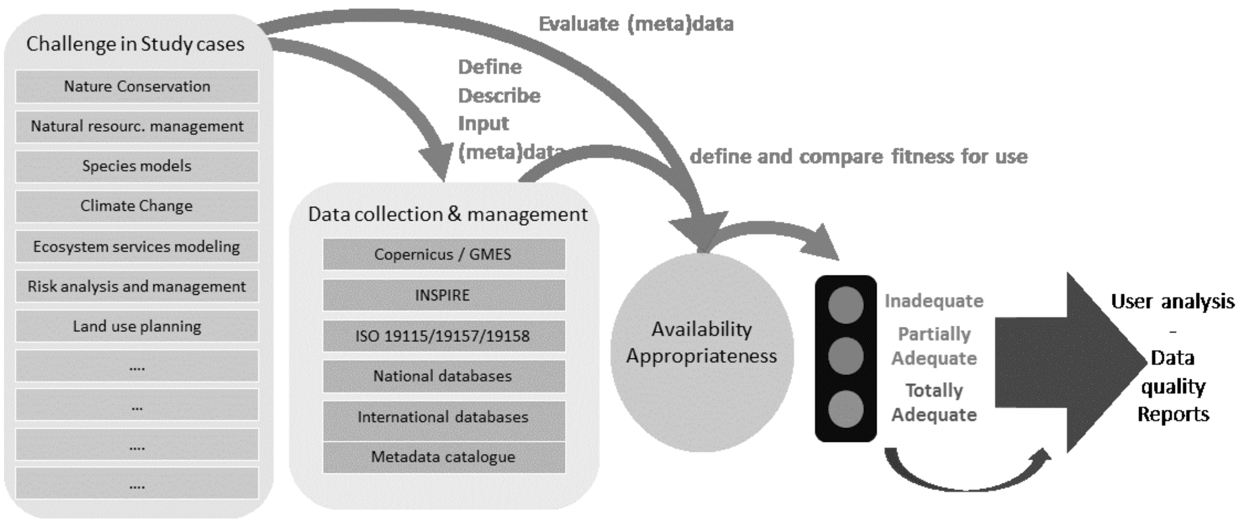


Fig. 2-6. (Meta)data quality assessment and fitness for use.

Collaborative (geo)processing models have become one of the major solutions to significantly enhance the capacity to share data/information, which are critical for the support of comprehensive analyses in a virtual geographic and knowledge network environment. With the emergence and growing maturity of the cloud computing infrastructure (Baranski et al., 2011), a cloud-based platform for collaborative (geo)processing models promises to provide a pattern for the next generation of (geo)processing collaboration (Evangelidis et al., 2014).

Spatial Data Quality Evaluation (SDQE) processes and tools aim at supporting assertive and effective spatial data (quality) management considering the implementation of Quality Control (QC) and Quality Assurance

(QA) procedures/tools in process, project and communities and knowledge networks environments. In spatial data quality assessment, ISO/TS 19158:2012 (Geographic Information – Quality assurance of data supply)[11] provide a framework for quality assurance specific to geographic information. It is based upon the quality principles and quality evaluation procedures of geographic information identified in ISO 19157 and the general quality management principles defined in ISO 9000. The framework defined in ISO/TS 19158:2012 (ISO, 2012) enables a customer to satisfy itself that its suppliers, both internal and external, are capable of delivering geographic information to the required quality. Fundamental to the framework is the assurance of the supplier's ability to understand and meet the quality requirements.

Through the quality assurance framework both the customer and the supplier are able to consider the quality required at the earliest opportunity in the production/update process (data time series) related to evaluate/perform quality of process. The SDQE implementation implies considering (spatial) information systems governance standards (ISO/IEC 38500:2008**,** *Corporate governance of information technology;* ISO 27001, the international information security standard) and total quality management (ISO 9000 series standards). The QC/QA implies developing, implementing, assessing and documenting procedures from human and organizational capacity building (individual and organizational capacity building (Human component at specification and implementation methods); e.g. ISO 19122:2004 - Qualification and certification of personnel https://www.iso.org/standard/31088.html); (meta)data collection, processing and sharing; information system/infrastructures technological development to define quality assessment management policies/strategies as well as efficient communication tools of QC/QA results (Fig. 2-7).

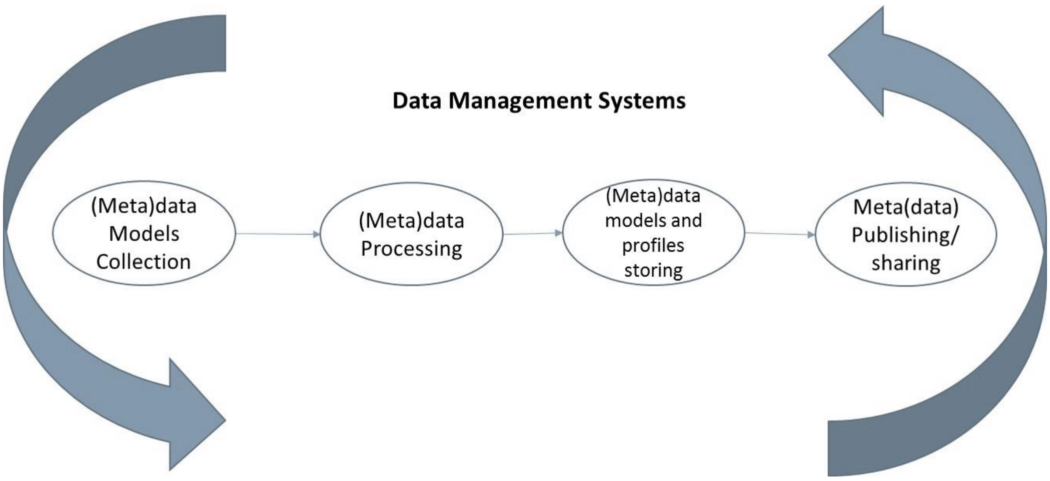


Fig. 2-7. Quality Assurance process for data management systems.

Data harmonization and implementation of data standards, metadata completion and sharing of metadata catalogue services are crucial aspects for data/systems interoperability. This stimulates the collaboration and promotes the results and the expected impacts (knowledge, communication and technical-political decision

[11] Principles and responsibilities of the relationship between the customer and the supplier that facilitate the framework are provided. The responsibility for the quality assessment procedure is shared between the customer and the supplier. ISO/TS 19158:2012 is applicable to customers and suppliers of all geographic information where the quality of the product may be impacted upon by the supplier's processes in any of the following scenarios: (1) there is an agreement or legislation for the supply of data acquisition services, (2) data acquisition services are being tendered for, and (3) one or more suppliers exist in the supply chain. ISO/TS 19158:2012 is not applicable for the supply of legacy datasets or "off the shelf' products where there is no further data production or update activity to manage.

support). Spatial data evaluation and management are critical in innovative, open, collaborative, multiuser/multifaceted and dynamic technical, scientific and political environments. In fact, the development of spatial data quality evaluation framework and virtual processing environment supports dynamic/circular and collaborative/distributed decision-making along projects, (multi)disciplinary communities and open/linked knowledge networks.

3. Identification, analysis and specification of metadata requirements and the user-oriented quality evaluation routines

To evaluate external data quality of spatial datasets based on metadata, it is necessary that metadata entries/catalogues contain information about the content, quality parameters, access and use conditions, and other attributes of datasets that can be used for external (meta)quality evaluation and also for knowledge discovery, indexing and searching. It is important to ensure that the metadata profile defined in other tasks will allow the application of user-oriented quality evaluation routines. This assessment was made in cooperation with Task 5.3, and new fields (quality elements) to be included in the DEIMS-SDR Dataset MD model were discussed in order to effectively allow quality evaluation based on metadata (ECOPOTENTIAL D5.2).

3.1 Spatial metadata profiles analysis and data quality assessment

The identification, analysis and specification of requirements for the user-oriented quality evaluation routines, supported on metadata, included:

(1) analysis of several candidate metadata profiles (INSPIRE; DEIMS-SDR; EML; ISO19157);

(2) evaluation of the current DEIMS-SDR Dataset MD model for support data quality assessment; and

(3) proposal of new fields for the selected metadata profile, related to data quality elements (in cooperation with WP 5.3).

Task 5.5 activities are focused on the user's input data quality requirements according to the application context (spatial modelling) and to the users' expertise (or expected, recognized and perceived quality). Therefore, an online questionnaire was developed (also in cooperation with Task 5.3) around what quality elements would be more relevant and useful for data providers and for data users.

3.1.1 Analysis of candidate metadata profiles

In the last ten years, several national and European projects have been able to accumulate metadata profiles and catalogues, methods, and new tools for evaluating spatial data quality. The collaborative networks management challenges reveal relevant scientific needs and associated opportunities in the metadata (semi)automatic generation and enrichment, expediting tools for requesting and evaluating the external quality of spatial data (GEOSS), as well as advances in the processes of control, assurance and quality management. The first step of activities related to "identification, analysis and specification of metadata requirements and the user-oriented quality evaluation routines" included the identification of candidate metadata profiles to support the development and the implementation of user-oriented quality evaluation routines.

Several candidate metadata profiles were analysed, namely:

- **DC (Dublin Core metadata)** is an open organization supporting innovation in metadata design and best practices across the metadata ecology[12]. The activities include work on architecture and

[12] See http://dublincore.org/

modelling, discussions and collaborative work. The Dublin Core Metadata Initiative (DCMI) supports shared innovation in metadata design and best practices across a broad range of purposes and business models. The DCMI's principles of operation are: open consensus building; international scope and participation; neutrality of purposes and business models; neutrality of technology; cross disciplinary focus. DCMI is therefore an organization committed to the development and open availability of resources that support a healthy global metadata ecosystem.

- **DwC (Darwin Core metadata)** is designed to facilitate the exchange of information about the geographic occurrence of organisms and the physical existence of biotic specimens in collections[13]. Extensions to the Darwin Core provide a mechanism to share additional information, which may be discipline-specific, or beyond the commonly agree upon scope of the Darwin Core itself. The Darwin Core and its extensions are minimally restrictive of information content by design, since doing so would render the standard useless for the implementation of data quality tools. On the Website of DwC we can find collections of any kind of biological objects or data; terminology associated with biological collection data; striving for compatibility with other biodiversity-related standards; and facilitating the addition of components and attributes of biological data.
- **TDWG (Biodiversity Information Standards)**, also known as the Taxonomic Databases Working Group, is a not for profit scientific and educational association that is affiliated with the International Union of Biological Sciences. TDWG was formed to establish international collaboration among biological database projects. TDWG promoted the wider and more effective dissemination of information about the World's heritage of biological organisms for the benefit of the world at large. Biodiversity Information Standards (TDWG)[14] now focuses on the development of standards for the exchange of biological/biodiversity data. The TDWG community's priority is the development of standards for the exchange of biological/biodiversity data. The most widely deployed formats for biodiversity occurrence data are Darwin Core and ABCD (Access to Biological Collection Data). The TDWG GUID Applicability Statement provides guidance on the assignment of Globally Unique Identifiers to biodiversity information resources. The mission of TDWG is to develop, adopt and promote standards and guidelines for the recording and exchange of data about organisms. Promote the use of standards through the most appropriate and effective means and act as a forum for discussion through holding meetings and through publications.
- **GEOMS (Generic Earth Observation Metadata Standard),** The GEOMS metadata and data structure requirements document may be applied to any project where data has to be exchanged. That outlines the metadata and data structure requirements developed to facilitate the use of geophysical datasets by improving their portability and accessibility, and by making their contents self-describing. This approach was originally selected to deal with atmospheric and oceanographic datasets, but has been recently expanded to support all measurements from Earth observation instruments. The GEOMS guidelines[15] describe the standard metadata definitions adopted for the correlative, experimental and model data archived for the EOS (Earth Observing System) - Aura validation program, the Envisat calibration and validation campaign, data from NDACC (Network for Detection of Atmospheric Composition Change), and the GECA project (Generic Embodied Conversational Agent), which supports existing and future ESA (European Space Agency) calibration and validation programs. A further development of the metadata standard previously implemented for the Envisat

[13] See http://www.tdwg.org/activities/darwincore/
[14] See http://www.tdwg.org/
[15] See https://evdc.esa.int/documentation/geoms/

Cal/Val activity, but the standard is now further generalized and harmonized to be implemented for EVDC (Atmospheric Validation Data Centre), AVDC (Aura Validation Data Centre) and NDACC. To facilitate the exchange of validation data among investigators and missions, data centers like the AVDC require that all validation data submitted to the archive are formatted using the Hierarchical Data Format (HDF) in conjunction with the GEOMS.

- **INSPIRE and MS Specification** is a European Community Directive, which entered into force in May 2007. The INSPIRE directive defines the guidelines for the establishment of a spatial data infrastructure in Europe in order to support the Community environmental policies, and policies or activities which may have an impact on the environment. The descriptive metadata are based on ISO19115/19139 as defined in the INSPIRE Metadata regulation (Regulation 1205/2008), and the minimum requirements expressed in the implementing rules have also to be met semantically, with metadata contents strictly satisfying the INSPIRE requirements. For the implementation, technical guidelines and implementing rules have been specified[16].
- **DEIMS (Dynamic Ecological Information Management System)** is used in order to collect information on characteristics of the protected areas, as well as data generated there. DEIMS is a web-based online editor for site and dataset metadata. The **ECOPOTENTIAL project** aims to provide information to analyse effects of global and regional threats, e.g. climate change and deposition, within protected areas along a geographic gradient. Data are provided by the protected areas themselves (in-situ data) as well as through remote sensing techniques (RS/EO data). The dataset metadata schemata describe the data files and data services provided by any data provider, based on an initial selection of metadata elements for long term data established within the *EnvEurope project*. The dataset metadata model is the compromise between the efforts for metadata provision, the needs for the data discovery and the requirements for the data re-usability. All selected metadata elements have been mapped to a corresponding metadata element in ISO19115/19139 and EML. Currently the community profile of the **Dataset MD Model** defined in **DEIMS-SDR** is adopted in ECOPOTENTIAL[17]. The community profile is based on metadata elements defined by the INSPIRE metadata specification and EML[18].

The development of spatially explicit models implies choosing and integrating adequate data and presenting results/effective proposals supported on data quality assessment and management principles/procedures. Data quality and data quality evaluation has received attention in geographic information science community (ISO/TC 211 Geographic information/Geomatics and OGC) with standards, technical guidance's and tools implementation. The last decades (in particular the last five years) big, open and linked data trends highlight the dynamics and need of formalizing spatial data quality orientations and standards. Recent ISO standards

[16] See http://inspire.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf
[17] A full description of the metadata elements can be found within the DEIMS-SDR data model documentation: https://data.lter-europe.net/deims/documentation/dataset
[18] **EML (Ecological MD Language)** is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts. The extensions of EML is the GBIF metadata profile that was established by a global network of countries and organizations. GBIF is a web portal promoting and facilitating the mobilization, access, discovery and use of biodiversity data. The portal uses a profile of EML; a How-to-Guide and Reference Guide for using the profile are available. The tools are integrated publishing toolkit, that it's a software platform using Darwin Core and EML to facilitate the efficient publishing of biodiversity data on the Internet, using the GBIF network. Metacat is a repository for data and metadata that helps scientists finding, understanding, and effectively using the data sets they manage or that have been created by others. Morpho is an application for accessing and manipulating metadata and data (both locally and on the network), with wizards creating metadata files using a subset of Ecological Metadata Language (EML)[18]. The uses case, which are Global Biodiversity Information Facility (GBIF); The Knowledge network for biocomplexity (KNB); Long Term Ecological Research Network; and National Center for Ecological Analysis and Synthesis.

data quality (ISO 19157:2013 and ISO19158) defines concepts and methods of spatial data quality assessment related to quality insurance of data supply, namely documentation of spatial data quality (schemas and guidelines):

- **ISO19115-1:2014 - Metadata Part1: Fundamentals,** defines the schema required for describing geographic information and services by means of metadata. It provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties. It is applicable to the cataloguing of all types of resources, clearinghouse activities, and the full description of datasets and services; geographic services, geographic datasets, datasets series, and individual geographic features and feature properties. ISO 19115-1:2014[19] is applicable to digital data and services, its principles can be extended to many other types of resources such as maps, charts, and textual documents as well as non-geographic data.
- **ISO 19115-2:2009 - Metadata Part 2: Extensions for imagery and gridded data**, ISO 19115-2:2009[20] extends the existing geographic metadata standard by defining the schema required for describing imagery and gridded data. It provides information about the properties of the measuring equipment used to acquire the data, the geometry of the measuring process employed by the equipment, and the production process used to digitize the raw data. This extension deals with metadata needed to describe the derivation of geographic information from raw data, including the properties of the measuring system, and the numerical methods and computational procedures used in the derivation. The metadata required to address coverage data in general is addressed sufficiently in the general part of ISO 19115.
- **ISO/TS 19139:2007 - Metadata - XML schema implementation,** defines Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from ISO 19115[21]. This technical specification is meant to enhance interoperability by providing a common specification for describing, validating and exchanging metadata about geographic datasets, dataset series, individual geographic features, feature attributes, feature types, feature properties, etc. ISO/TS 19139:2007 included: (i) requirements (e.g. rule-based; quality; rules for application schema); (ii) extensions to the UML models in the ISO19100 series of International Standards; (iii) encoding rules; and (iv) encoding descriptions (e.g. gmd namespaces).
- **ISO 19157:2013 - Data quality**, replaces **ISO/TS 19138:2006 Data quality measures[22]**, **ISO 19114:2003 Quality evaluation procedures[23]**, and **ISO 19113:2002 Quality principles[24]**, and establishes the principles for describing the quality of geographic data. This standard (i) defines components for describing data quality; (ii) specifies components and content structure of a register for data quality measures; (iii) describes general procedures for evaluating the quality geographic data; (iv) provides guidelines on how to describe, evaluate and report data quality. ISO 19157:2013[25] also defines a set of data quality measures for use in evaluating and reporting data quality. It is applicable to data producers providing quality information to describe and assess how well a data set conforms to its product specification and to data users attempting to determine whether or not

---

[19] See https://www.iso.org/standard/53798.html
[20] See https://www.iso.org/standard/39229.html
[21] See https://www.iso.org/standard/32557.html
[22] See https://www.iso.org/standard/32556.html
[23] See https://www.iso.org/standard/26019.html
[24] See https://www.iso.org/standard/26018.html
[25] See https://www.iso.org/standard/32575.html

specific geographic data are of sufficient quality for their particular application.

- **ISO/TS 19157-2:2016 - Data quality - Part 2: XML schema implementation**[26], defines data quality encoding in XML. This technical specification is an XML schema implementation derived from ISO 19157:2013 and the data quality related concepts from ISO 19115-2:2009 “Metadata - Part2: Extensions for imagery and gridded data”. ISO/TS 19157-2:2016 utilizes encoding rules from ISO 19118:2011 “Encoding” and ISO/TS 19139:2007 “Metadata - XML schema implementation”, and the implementation approach from ISO/TS 19115-3:2016 “Metadata - Part3: XML schema implementation for fundamental concepts” to define an XML schema implementation of ISO 19157:2013, and the data quality related concepts from ISO 19115-2:2009.
- **ISO/TS 19158:2012 - Quality assurance of data supply**, provides a framework for quality assurance specific to geographic information. This technical specification is based upon the quality principles and quality evaluation procedures of geographic information identified in ISO 19157 and the general quality management principles defined in ISO 9000. The framework defined in ISO/TS 19158:2012[27] enables a customer to satisfy itself that its suppliers, both internal and external, are capable of delivering geographic information to the required quality. Fundamental to the framework is the assurance of the supplier's ability to understand and meet the quality requirements. Through the quality assurance framework both the customer and the supplier are able to consider the quality required at the earliest opportunity in the production/update process. Principles and responsibilities of the relationship between the customer and the supplier that facilitate the framework are provided. The responsibility for the quality assessment procedure is shared between the customer and the supplier. ISO/TS 19158:2012 is applicable to customers and suppliers of all geographic information where the quality of the product may be impacted upon by the supplier's processes in any of the following scenarios: (i) there is an agreement or legislation for the supply of data acquisition services, (ii) data acquisition services are being tendered for, and (iii) one or more suppliers exist in the supply chain.

The scientific and technological community of (spatial) data quality assessment presents new approaches conceptual and methodological frameworks, methods and instruments/tools in the direct (total or partial) spatial data evaluation (ISO 19157:2013) namely, in the generation of indicators and representation of the spatial variability of quality elements. This requires simple and (semi)automatic metadata fulfilment, integration and metadata catalogue interoperability (ISO 19115-1:2014 and ISO/TS 19139:2007), as well as the need of operating theoretical methods foreseen in the spatial data quality assessment and management standards (ISO 19157:2013 and ISO 19158:2012). The ISO standards indicate data quality elements introduced on used principal metadata profile.

To evaluate external data quality of spatial **datasets** and **datasets series** based on metadata, it is necessary that metadata entries contain fields about the content, quality parameters, access and use conditions, and other characteristics of datasets that can be used for external (meta) quality evaluation and also for knowledge discovery, indexing and searching.

As such, this assessment was made in cooperation with task 5.3, and new fields (data quality sub-elements) to be included in-situ and other spatial metadata profile were discussed in order to effectively allow quality

[26] See https://www.iso.org/standard/66197.html

[27] See https://www.iso.org/standard/32576.html

evaluation based on metadata.

A comparative analysis of different metadata profiles indicates insufficiency and omissions related to data quality aware elements fields (Tab. 3-1). The conformance of an ISO 19115 metadata set to the ISO 19115 (mandatory elements) does not guarantee the conformance to INSPIRE. The use of these guidelines to create INSPIRE metadata ensures that the metadata is not in conflict with ISO 19115[28]. The requirements INSPIRE for spatial dataset and dataset series (implementing rules for metadata: regulation 1205/2008, and technical specifications ISO 19139), for some metadata elements, are more demanding than ISO 19115 and DEIMS-SDR MD (see Tab. 3-1), namely: (i) "geographic bounding box", INSPIRE is more restrictive (bounding box is mandatory); (ii) "temporal extent", INSPIRE is more demanding, as a temporal reference is mandatory, and can be expressed as a temporal extent; (iii) "lineage", INSPIRE is more demanding (a general lineage statement is mandatory); in case of DEIMS-SDR the fields: "Dataset Methods description", "Dataset Instrumentation", "Dataset Sampling description" and "Quality assurance" may be correspond to the "lineage"; (iv) "responsible organization", INSPIRE is more demanding by requiring as mandatory both the name of the organization, and a contact e-mail address; (v) "metadata point of contact", INSPIRE demands the name of the organization, and a contact e-mail address.

INSPIRE is more demanding, because it includes the fields: Resource type, Unique Resource Identifier, Keyword, Conditions for access and use, Limitations on public access, and Conformity (not included in ISO). However, ISO 19115 includes some relevant optional fields not found in regulation 1205/2008, namely: "distribution format", "spatial reference system", "reference system", and also some fields related metadata description/date of creation or revision/versioning, but recommended/mandatory for compliance with regulation 1089/2010 (implementing rules for interoperability of spatial datasets).

Tab. 3-1. Description and nature of fields included in ISO 19115, requirements of INSPIRE/ISO 19139, DEIMS-SDR Dataset MD and ISO 191157 metadata standards (Legend: Mandatory- [M]; Conditional- [C]; Optional- [O]).

Metadata Element	Description	Type of field	ISO 19115	INSPIRE /ISO 19139	*DEIMS-SDR MD*	ISO 19157
Metadata	This is the description of the organisation responsible for the creation and maintenance of the metadata. The date which specifies when the metadata record was created or updated. This is the language in which the metadata elements are expressed.	Text and Date	[M]	[M]	*[M]*	---
Resource title	This is a characteristic, and often unique, name by which the resource is known. This field refers to the title of a specific dataset [e.g. a dataset of distribution information for the population of bats should be referred as "bats distribution data"]. The titles should be short (in length) and objective.	Text	[M]	[M]	*[M]*	---
Resource abstract	This is a brief narrative summary of the content of the resource with no more than 200 characters.	Text	[M]	[M]	*[M]*	---
Resource type	Scope to which metadata applies "MD_ScopeCode" (Data type – see annex B.5.25 of ISO 19115).	List	---	[M]	---	[M]
Resource locator	Location (address) for on-line access using a Uniform Resource Locator address or similar addressing scheme.	URL or *List*	[O]	[C]	*[O]*	---
Unique resource identifier	Value uniquely identifying an object within a namespace "MD_Identifier".	URI	---	[M]	*[M]*	---
Topic category	The topic category is a high-level classification scheme to assist in the grouping and topic-based search of available spatial data resources "MD_TopicCategory", according to the ISO 19115.	List	[M]	[M]	*[M]*	---
Keywords	Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject. Note: related to identification of the INSPIRE	Text or	---	[M]	*[M]*	---

[28] However, full conformance to ISO 19115 implies the provision of additional metadata elements which are not required by the INSPIRE Implementing Rule on Metadata.

Metadata Element	Description	Type of field	ISO 19115	INSPIRE /ISO 19139	*DEIMS-SDR MD*	ISO 19157
	Spatial Data themes (thematic category) according to Directive INSPIRE. The keyword set is related keywords describing the content of the dataset derived from the controlled vocabulary implemented by EnvThes-thesaurus for long term ecological research, monitoring, experiments EnvThes and other environmentally related thesauri as EUNIS Habitats and others.	*List*				
Spatial resolution	Spatial resolution refers to the level of spatial detail of the dataset. It shall be expressed as a set from zero to many resolution distances (typically for gridded data and imagery-derived products) or equivalent scales (typically for maps or map-derived products). An equivalent scale is generally expressed as an integer value expressing the scale denominator. A resolution distance shall be expressed as a numerical value associated with a unit of length.	Numeric	[O]	[M]	*[M]*	---
Temporal extent	The temporal extent defines the time period covered by the content of the resource.	Date	[O]	[M]	*[M]*	---
Date of publication	This is the date of publication of the resource when available, or the date of entry into force.	Date	[M]	[M]	*[O]*	---
Geographic bounding box	This field refers to the geographical scope of the dataset, particularly whether the dataset covers all or just a portion of the study area. 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.	Numeric	[C]	[M]	*[M]*	---
File format	This field refers to the type of file of the dataset (distribution format).	Text	[O]	---	---	---
Author	This field refers to the institution or individual that produced the dataset (Responsible organisation).	List	[M]	[M]	*[O]*	---
Property	This field refers to the property of the dataset being necessary to state if there are any conditions applying to its access and use.	List	---	[M]	*[O]*	---
Spatial Reference System	This field refers to the geographical reference system of the dataset.	List	[O]	---	---	---
Lineage	Lineage describes the history of a dataset and recount the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form. This general, non-quantitative information is illustrative for users and can help assessing the quality of a dataset, especially in cases where it is used for a particular application that differs from the intended application.	Text	[O]	[M]	*[M]*	---
Maintenance information	This field refers to the information about the scope and frequency of updating.	Text	---	[O]	---	---
Data quality	Quality evaluation processes are used **in different phases of a product life cycle, having different objectives in each phase.** The process for evaluating data quality is a sequence of steps to produce a data quality result, namely: (i) **data quality unit(s) identification**: composed of a **scope** and **quality element(s),** all data quality elements relevant to the data for which quality is to be described should be used; (ii) **data quality measures identification**: if applicable a measure should be specified for each data quality element; (iii) **data quality evaluation procedures selection**: a data quality evaluation procedure consists of applying one or more evaluation methods; (iv) **determine the output of the data quality evaluation** (quantitative, qualitative, conformance): this result is the output of applying the evaluation; (v) **reporting data quality** (metadata). **Data quality evaluation process** (see ISO/TS19157-2 XML)**:** (1) **Description of data quality unit(s)**: **Scope** [MD_Scope]: ISO19115-1 (2) **Data quality evaluation element(s)**: [DQ_Element(s)]*see Tab. 3-2 (3) **Data quality measures:** [DQM_Description_DataType]; (MD_Identifier_DataType ISO19115-1); (4) **Data quality evaluation method:** [EvaluationMethodType]; CodeList: directInternal; directExternal; indirect.	Text or List or Numeric	--- (Data quality was moved to ISO19157	[O]	---	[O]

Metadata Element	Description	Type of field	ISO 19115	INSPIRE /ISO 19139	*DEIMS-SDR MD*	ISO 19157
	(5) **Result** (Meta)quality elements are a set of quantitative and qualitative statements about a quality evaluation and its result. The knowledge about the quality and the suitability of the evaluation method, the measure applied and the given result may be of the same importance as the result itself. **Data quality result (metaquality)** (Meta)quality may be described using the following elements (report): (1) Confidence; (2) Representativity; (3) Homogeneity.					
Quality assurance	Quality assurance (QA), is defined with "all the planned and systematic activities implemented within the quality system, and demonstrated as needed, to provide adequate confidence that as entity will fulfil requirements for quality". This analysis is done according to a requisition (product, legal or normative obligation, ...). Recommendations: Several spatial data producers normally have an obligation of mean but no obligation of result. Consequently, QA is not sufficient since it cannot guarantee the production of quality data.	List	---	---	*[O]*	--- *(ISO/TS 19158)*

Tab. 3-1 (cont.)

Compliance data quality elements and evaluation process (described in metadata) need the definition of scope[29] as the first and critical step for data quality evaluation process, since this is linked to others stages of the procedure, namely: selection of data quality elements and sub-elements (DQ_Element(s)) (for each DQ_SubElement selected to be evaluated, see Tab. 3-2), data quality measures (DQM_Description_DataType), data quality evaluation method (EvaluationMethodType), and expected result (Result). The result is the output of applying the evaluation and should be reported through metadata (Reporting).

The metadata comparative analysis indicates the importance of increasing and detailing the quality elements fields related to descriptors of quality elements according to ISO19157:2013, in order to improve and guarantee quality control (and report by metadata) in the production and maintenance (scope and frequency of updating) of quality throughout the data life cycle (data management) (Tab. 3-2).

The analysis (Tab. 3-1 and Tab. 3-2) indicates the importance and possibility of suggesting/adding new data quality (sub)elements (recommended by ISO19157) into the metadata profiles in order to complement the missing elements as well as to value/support the data quality evaluation exercises/processes.

Tab. 3-2. Examples of data quality elements, sub-elements and measures (according to ISO19157:2013).

DQ_Element(s)	Examples DQ_SubElement(s)	Examples Measures (DQM_Description_Datatype)
Completeness	Completeness commission	Number of excess items/ Integer/ error count
	Completeness omission	Number of missing items/ Integer/ error count
Logical Consistency	Conceptual consistency	Number of items not compliant with the rules of the conceptual schema/ Integer/ error count
	Domain consistency	Number of items not in conformance with their value domain/ Integer/ error count
	Format consistency	Number of physical structure conflicts/ Integer/ error count
	Topological Consistency	*Select measure*
	Vertical positional uncertainties	Root mean square error (RMSE)/ Measure

[29] The definition of scope [MD_Scope] according to ISO19115-1:2013 is defined by a list reference: (a) a **dataset series**; (b) a **dataset**; and (c) a **subset of data defined by one or more of the following characteristics**: (i) types of items (sets of feature types, feature attributes, feature operations or feature relationships); (ii) specific items (sets of feature instances, attribute values or instances of feature relationships); (iii) geographic extent; (iv) temporal extent (the time frame of reference and accuracy of the time frame).

DQ_Element(s)	Examples DQ_SubElement(s)	Examples Measures (DQM_Description_Datatype)
Positional Accuracy	Horizontal positional uncertainties	Root mean square error of planimetry (RMSEP)/ Measure
	Gridded data positional accuracy	*Select measure*
Thematic Accuracy	Thematic classification correctness	Incorrectly classified features/ %/ error indicator
		Misclassification matrix/ Integer – Matrix (n*n)
		Kappa coefficient/ Integer
	Non quantitative attribute correctness	Number of incorrect attribute values/ Integer/ error count
	Quantitative attribute accuracy	Attribute value uncertainty at 95 % significance level
Temporal Quality	Accuracy of a time measurement	Time accuracy at 50% significance level (interval defined by an upper and a lower limit)/ Measure
	Temporal consistency	Chronological error/ Boolean/ error indicator
	Temporal validity	*Select measure*
Usability	Usability element	Degree of adherence of a dataset to a specific set of requirements (specifications)

Tab. 3-2 (cont.)

3.1.2 Metadata evaluation and data quality assessment

The purpose of describing the quality of geographic data is to facilitate the comparison and selecting of the dataset best suited to application user´s needs or requirements. Information on the quality of geographic data allows a data producer to evaluate how well a dataset meets the criteria set forth in its product specification and assists data users in evaluating a product’s ability to satisfy the requirements for different contexts/applications. The purpose of describing the quality of geographic data is to facilitate the comparison and selecting of the dataset best suited to application needs or requirements as well as to reduce possible data degradation across time (Tessarolo et al., 2017).

Data producers are interested in the widest reuse datasets that is only possible when - the data sets satisfy the requirements of the targeted users and the degree to which the requirements are fulfilled is documented. User’s requirements specific targets on data quality have to be achieved in the course of data production or transformation. The second aspect corresponds to documenting the quality of the data that is eventually delivered to the users. For each of these tasks a common way of expression is necessary that comprises an agreed terminology, evaluation, and reporting methods, both aspects lead to formalising data quality.

In order to implement data quality evaluation routines for geographic data (Bédard and Vallière, 1995) proposed six characteristics (criterion applied, see column “Main Characteristics” on Tab. 3-3) to define the external quality of a geospatial dataset: (i) **definition**: to evaluate whether the exact nature of a data and the object that it describes, that is, the “what”, corresponds to user needs (semantic, spatial and temporal definitions); (ii) **coverage**: to evaluate whether the territory and the period for which the data exists, that is, the “where” and the “when”, meet user needs; (iii) **lineage**: to find out where data come from, their acquisition objectives, the methods used to obtain them, that is, the “how” and the “why”, and to see whether the data meet user needs; (iv) **precision**: to evaluate what data is worth and whether it is acceptable for an expressed need (semantic, temporal, and spatial precision of the object and its attributes); (v) **legitimacy**: to evaluate the official recognition and the legal scope of data and whether they meet the needs of de facto standards, respect recognized standards, have legal or administrative recognition by an official body, or legal guarantee by a supplier, etc.; and (vi) **accessibility**: to evaluate the ease with which the user can obtain the data analysed (cost, time frame, format, confidentiality, respect of recognized standards, copyright, etc.).

To evaluate external data quality of spatial datasets based on metadata, it is necessary that metadata entries contain information that can be used for external (meta)quality evaluation and also for knowledge discovery, indexing and searching. As such, efforts have been made to ensure that the adopted metadata standard/profile: **Dynamic Ecological Information Management System - Dataset MD Model** (DEIMS-SDR MD) will allow the application of user-oriented quality evaluation routines.

The documentation of datasets by all DEIMS-SDR MD standard metadata elements, provided that full and duly completed, allows inferring some quality elements (indicators). If in the documentation of datasets through this metadata standard only the required metadata elements were used (see Mandatory elements on Tab. 3-3), the possibilities of applying quality indicators supported in metadata is limited.

However, we present a proposal that, to the level of implementation of data quality evaluation routines, that allows applying of "General workflow for external quality evaluation framework - Pôças et al., 2014", including ISO19157:2013 implementations rules (see *Italic/grey* text in the Tab. 3-3). Data quality elements are implicit to DEIMS-SDR MD according possibilities to explore potential data quality evaluation but this metadata profile limits explicit data quality assessment exercises.

Data on the quality of geographic data allows a data producer to evaluate how well a dataset meets the criteria set forth in its product specification and assists data users in evaluating a product's ability to satisfy the requirements for their application. We recommend the additional of new fields (quality elements) to be included in the DEIMS-SDR MD [Data quality elements], or the adaptation of some profile fields, namely its structure and content [17. Dataset methods; 18. Dataset Instrumentation description; and 19. Dataset Sampling description], and whether it is mandatory or optional [10. Dataset Access and Use constraints; Dataset Intellectual Rights, and 12. Dataset Online distribution], in order to improve the possibilities for assessing the quality of spatial data.

Tab. 3-3. Cross-analysis/evaluation of the current metadata profile (DEIMS-SDR MD) in terms of adequacy for data quality evaluation framework.

Metadata contents (DEIMS)	**Element (DEIMS)**	***Main Characteristics***	***Examples of quality indicators***	***Description/Value domain/comparison type***
1. Dataset Title [Required: mandatory]	**1.1 Title** Free text	*Definition*	*Typology*	*Topic category defining the main data set theme/ Text string/ String comparison*
2. Dataset Identifier [Required: mandatory]	**2.1 Site Name** Text; Reference List [Content Type] SITE	*Definition*	*Typology Legitimacy [case study – PAs]*	*Dataset site/ Text string/ String comparison*
3. Dataset Creator and Contact Points [Required: mandatory]	**3.1 Dataset Owner/Creator** **3.2 Dataset contact point** Text; Reference to [Content Type] PERSON			
4. Dataset Metadata Provider [Required: mandatory]	**4.1 Metadata provider** Text; Reference to [Content Type] PERSON			*Dataset metadata (provider) characteristics/ Text string/ String comparison*
5. Dataset Metadata date [Required: mandatory]	**5.1 Date** Date as YYYY-MM-DD			
6. Dataset Publication Date [Required: optional]	**6.1 Date of publication** Date as YYYY-MM-DD			
7. Dataset Language [Required: optional]	**7.1 Language** Reference (LOV) based on **ISO 639 standardized nomenclature** used to classify all known languages			

Metadata contents (DEIMS)	Element (DEIMS)	*Main Characteristics*	*Examples of quality indicators*	*Description/Value domain/comparison type*
8. Dataset Abstract [Required: mandatory]	**8.1 Abstract** Free text	*Definition*	*Typology*	*Abstract defining the main characteristics the dataset/ Text string/ String comparison*
9. Dataset Keyword set [Required: EnvThes Keywords - mandatory] [Required: Free Keywords - optional]	**9.1 EnvThes Keywords** Reference to [Taxonomy] **LTER Controlled Vocabulary**, which is regularly updated **9.2 Free Keywords** Free text	*Definition*	*Typology*	*Topic category defining the main data set theme/ Text string/ String comparison*
			Keyword	*Keyword defining the main data set theme/ Text string/ String comparison*
10. Dataset Access and Use constraints [Required: optional]	**10.1 Principal and granted permission** Reference List: Use constraints are defined for the following **User Groups**: *Administration; *Education & Training; *Public; *Research; *LTER-Europe; *Others Use constraints are defined by the following **Permissions**: *Free for access; *Free for access and use upon request; *Other restrictions according to rules defined in intellectual rights; *Restricted access detaily defined in intellectual property information; *No access	*Accessibility*	*Access and use constraints*	*Conditions applying to access and use (Reference List: **User Groups** and Reference List: **Permissions**)/ Text string/ String comparison*
11. Dataset Intellectual Rights [Required: optional] [Multiplicity: [0..n] Metadata is optional (provided if necessary)]	**11.1 Intellectual rights** Predefined list of IPR statements, and Other IPR for user defined free text statements. Reference List: *Co-authorship on publications resulting from use of the dataset; *The data provider must be offered co-authorship for publications using this dataset at least within the metadata description; *Formal acknowledgement of the dataset providers; *The opportunity to collaborate on the project using the dataset; *At least part of the costs of dataset acquisition, retrieval or provision must be recovered; *The opportunity to review the results based on the dataset; *Reprints of articles using the dataset must be provided to the data provider; *The dataset provider is given a complete list of all products that make use of the dataset; *Legal permission for dataset use is obtained; *Mutual agreement on reciprocal sharing of data; *The data provider is given and agrees to a statement of uses to which the dataset will be put; *Other rights	*Legitimacy*	*Producer recognition*	*Data producer recognition/ Text string/ String comparison*
			Intellectual Property Rights (IPR), copyright, and various property rights	*Intellectual Property Rights (IPR) for the dataset (Reference List)/ Text string/ String comparison*
			Intellectual Rights provide or reference a URL (web address)	*Intellectual Rights reference a URL/ Text string/ String comparison*
12. Dataset Online distribution [Required: optional] [Multiplicity: [0..n] for Dataset locator; Metadata is conditional (shall be provided if not Dataset file element provided)]	12.1 Online Locator **12.1.1 Distribution function** 12.1.2 Distribution URL **12.1.2.1 Web adress Title** **12.1.2.2 Web adress URL** **12.1.3 Email** Distribution function: Text (Reference list) Web address title Text (255) Web address URL Valid URL E-Mail Valid email address 12.2 WMS Related **12.2.1 WMS Map** 12.2.2 WMS Map Web Address **12.2.2.1 Web adress Title**	*Accessibility*	*Access and use constraints*	*Conditions applying to access and use the dataset/ Text string/ String comparison*
			File format	*Distribution file format/ Text string/ String comparison*
			Online distribution function	*Web address function (Reference List)/ Text string/ String*

Metadata contents (DEIMS)	**Element (DEIMS)**	***Main Characteristics***	***Examples of quality indicators***	***Description/Value domain/comparison type***
	12.2.2.2 Web adress URL			
13. Data Sources [Required: optional]	**13.1 Name** **13.2 Description** 13.3 Source **13.3.1 File upload** **13.3.2 Header Lines** **13.3.3 Footer Lines** **13.3.4 Orientation** **13.3.5 Quote character** **13.3.6 Field delimiter** **13.3.7 Record delimiter**	*Lineage*	*File structure*	*Describing the structure of the file format/ Text string or conceptual schema/ String comparison*
14. Geographic (Research site) [Required: mandatory]	Entity type Research site as geographic reference **14.1 Name of Research Location** **14.2 Description** **14.3 Research Location ID** **14.4 Related Site** 14.5 Geographic Location **14.5.1 North bound coordinate** **14.5.2 South bound coordinate** **14.5.3 West bound coordinate** **14.5.4 East bound coordinate** **14.5.5 Maximum Altitude** **14.5.6 Minimum Altitude** **14.6 Images** **14.7 Details** Reference to [Content Type] Research Site	*Coverage*	*Spatial extent*	*Bounding box defining spatial coverage/ Geometric boundary box/ % of cover*
15. Dataset Temporal extent [Required: mandatory]	**15.1 From date** **15.2 To date** Date	*Coverage*	*Temporal extent*	*Time interval defining temporal coverage/ Start date – End date/ % of cover*
16. Dataset Taxonomic coverage [Required: optional] [Multiplicity: [0..n] Metadata is conditional (shall be provided for biotic datasets)]	**16.1 Biological Classification** Reference to [Taxonomy] Biological Classification	*Definition*	*Keyword*	*Keyword defining the main dataset theme/ Text string/ String comparison*
17. Dataset Methods description [Required: mandatory]	17.1 Method Online Reference **17.1.1 Web address title** **17.1.2 Web address URL** **17.2 Description** Web address title Text (255) Web address URL Valid URL Description Text Reference List: Method and related concepts available in EnvThes	*Lineage*	*Lineage description in metadata*	*Description of data production methods, quality assurance and control/ Text string/ Boolean comparison*
18. Dataset Instrumentation description [Required: optional] [Multiplicity: [0..n] Metadata is mandatory]	**18.1 Instrumentation** Free text	*Lineage*	*Lineage description in metadata*	*Description of data production methods, quality assurance and control/ Text string/ Boolean comparison*
19. Dataset Sampling description [Required: mandatory] [Multiplicity: [1..n] for Spatial scale - metadata is mandatory; [1] for Sampling time span - metadata is mandatory;	19.1 Representative area of sampling **19.1.1 Spatial scale** 19.2 Sampling frequency **19.2.1 Sampling time span** **19.2.2 Minimum sampling unit** Predefined values defined in ECOPAR. Free text for option Other.	*Precision* *Lineage*	*Spatial scale* *Lineage description in metadata*	*Equivalent scale or spatial resolution defining the level of detail/ Integer-Double/ Intersection test* *Description of data production methods and procedures (sampling; …)*

Metadata contents (DEIMS)	Element (DEIMS)	*Main Characteristics*	*Examples of quality indicators*	*Description/Value domain/comparison type*
[1] for Minimum sampling unit - metadata is mandatory]				
20. Quality Assurance [Required: optional]	**20.1 Quality assurance** Free text *(see ISO19158:2012)*	*Lineage*	*Lineage description in metadata*	*Description of data quality assurance and control methods and procedures*
21. Dataset Legal obligation reporting [Multiplicity: [0..n] Metadata is optional]	**21.1 Legal act** Predefined list of relevant EU directives. Free text for option Other. Reference list: *Habitats Directive (92/43/EEC); *Water Framework Directive (00/60/EEC); *Bird Directive (79/409/EEC); *Marine Strategy Framework Directive; *Water Policy Directive; *None; *Other directive	***Legitimacy***	***Legal act***	*Description of data obligation report/ Text string (Reference List)/ String comparison*
Data quality [proposal based on ISO19157:2013]	***22.1 Data quality evaluation process*** *Description of data quality unit(s)* ***22.1.1 Scope*** *[MD_Scope] ISO19115-1:2014; List Reference:* *a) a dataset series;* *b) a dataset;* *c) a subset of data defined by one or more of the following characteristics:* *1) types of items (sets of feature types, feature attributes, feature operations or feature relationships);* *2) specific items (sets of feature instances, attribute values or instances of feature relationships);* *3) geographic extent;* *4) temporal extent (the time frame of reference and accuracy of the time frame).* ***22.1.2. Data quality evaluation element(s)*** *[DQ_Element(s)]* ***22.1.3 Data quality measures*** *[DQM_Description_DataType] (MD_Identifier_DataType ISO19115-1:2013);* ***22.1.4 Data quality evaluation method*** *[EvaluationMethodType] CodeList: directInternal; directExternal; indirect.* ***22.1.5 Data quality result (metaquality)***	***DQ_Element(s)*[30]**	***Examples DQ_Element(s)***	***Examples Measures (DQM_Description_Datatype)***
		Completeness	*Completeness commission*	*Number of excess items/ Integer/ error count*
			Completeness omission	*Number of missing items/ Integer/ error count*
			…	…
		Consistency	*Conceptual consistency*	*Number of items not compliant with the rules of the conceptual schema/ Integer/ error count*
			Domain consistency	*Number of items not in conformance with their value domain/ Integer/ error count*
			Format consistency	*Number of physical structure conflicts/ Integer/ error count*
			…	…
		Positional Accuracy	*Vertical positional uncertainties*	*Root mean square error (RMSE)/ Measure*
			Horizontal positional uncertainties	*Root mean square error of planimetry (RMSEP)/ Measure*
			Gridded Data Positional Accuracy	*Select measure*
			…	…
		Thematic Accuracy	*Thematic classification correctness*	*Incorrectly classified features/ %/ error indicator*
				*Misclassification matrix/ Integer – Matrix (n*n)*
				Kappa coefficient/ Integer
			…	…
		Temporal Quality	*Accuracy of a time measurement*	*Time accuracy at 50% significance level (interval defined by an upper and a lower limit)/ Measure*

[30] The selection the data quality evaluation element(s) and measure(s) more relevant must be defined according data typology (e.g. pre-existing and new reference and thematic n-situ data; EO data; Ancillary data and new modelled data …).

Metadata contents (DEIMS)	Element (DEIMS)	*Main Characteristics*	*Examples of quality indicators*	*Description/Value domain/comparison type*
			Temporal consistency	*Chronological error/ Boolean/ error indicator*
			…	*…*
		Usability	*Usability element*	*Degree of adherence of a dataset to a specific set of requirements (specifications)*
	22.2 (Meta)quality *22.2.1 Confidence* *22.2.2 Representativity* *22.2.3 Homogeneity*			

Table 3-3 (cont.)

3.2 Users' knowledge, interests, experiences and perceived utility of spatial data quality

Task 5.5 activities are focused on the user's input data quality requirements according to the application context (spatial modelling) and to the users' expertise (or perceived, recognized or expected quality). Therefore, an online questionnaire was developed (also in cooperation with Task 5.3 and other WP5 tasks) around what quality elements would be more useful/important for data providers and for data users. An online questionnaire "knowledge and routines of data quality assessment and management" was devised to inquire internal ECOPOTENTIAL researcher's/data user's and is targeted at both data providers and data users (including EO data production and usage; field data production and usage; application of data in correlative or process-based modelling), but also at other users such as technical and political decision-makers (e.g. PA managers).

The questionnaire on "knowledge and routines of data quality assessment and management" aims to inquire the ECOPOTENTIAL community of data providers and data users about their knowledge, their practical experience and their awareness (utility) of data quality evaluation routines/tools (Annex I). The questionnaire can be found and answered online[31]. It consists of a collaborative and oriented online questionnaire with closed and open questions, disseminated to the ECOPOTENTIAL community by the Basecamp platform and mailing list (by the WP5 team Leader) (Fig. 3-1). The questionnaire is aimed to inquire individual ECOPOTENTIAL researchers, including data providers and data users, on quality issues related to EO data production and usage, field data production and usage, and application of data ecological models. More specifically: (1) knowledge of data quality, including quality theory, concepts, elements (spatial data user´s theoretical knowledge); (2) current practices (e.g. practical experience) related to (spatial) data quality assessment and management routines (spatial data user´s practical experience); and (3) awareness, interest and willingness to implement data quality routines (spatial data user´s perceived utility).

The questionnaire results (Annex II) will be relevant for future activities in WP5, namely when: (1) defining the fields/attributes of quality elements to include in the selected metadata profile (Task 5.3); (2) specifying and developing methods (external evaluation) and tools (routines and information/ technological application) of spatial data quality assessment (Task 5.5); and (3) devising proposals for implementation of

[31]The questionnaire can be found and answered at:
https://docs.google.com/forms/d/e/1FAIpQLScQthxJtUTnS74xJbU-5dSoGlQeWUEJKplQ4W0rRIhx1dyLTQ/viewform

quality management processes (WP5). The estimated time to respond the questionnaire was approximately 10/15 minutes, and three ECOPOTENTIAL internal calls were made in order to obtain the greatest number of possible responses (first call until 31st March 2017; second reminder 14th April 2017; and third reminder 31st May 2017).

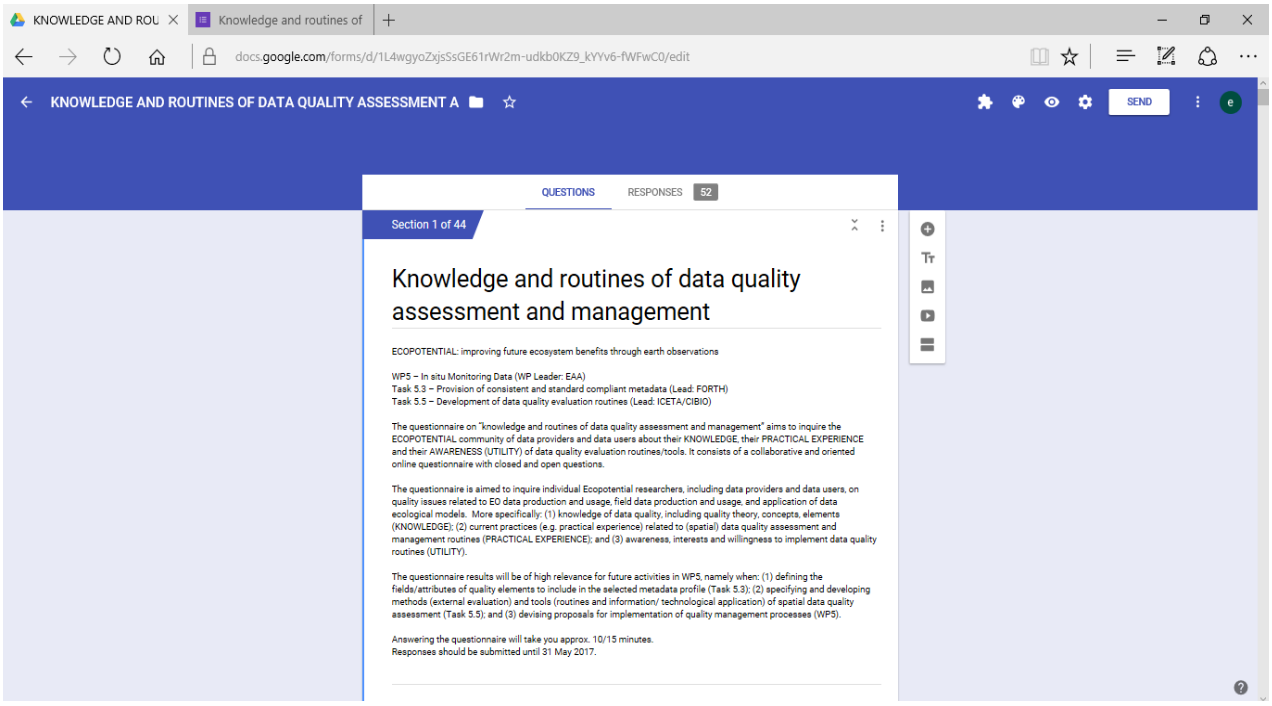


Fig. 3-1. Implementation through Google Forms of the questionnaire on "knowledge and routines of data quality assessment and management".

3.2.1 ECOPOTENTIAL spatial data users and their position within the spatial data life cycle

The purpose of first question was to identify the user's (ECOPOTENTIAL community):

- **academic background**: Six fixed answers and one open on have been selected. The possible answers are: () Biology/Ecology; () Mathematics/Statistics; () Engineering/Technology; () Geophysics/ Geography; () Human and Social Sciences; () Remote Sensing/Spatial data; () Other (open).
- **academic degrees and institution:** Four fixed answers and one open on have been selected. The possible answers are: () Post PhD; () PhD; () Master; () Degree; () Other (open).
- **group/tasks in which** the participant is involved. **Partner/Country,** the possible answers are: Italy (CNR, UNILE, EURAC, ISPRA, POLIMI); Spain (CSIC, CREAF, UAB, IISTA-UGR, STARLAB, REDIAM); Germany (UFZ, KIT, UBT, DLR, UPotsdam, MfN, iDiv-MLU); France (CNRS-UMR, TdV, CESBIO-UPS, UBO); UK (UNIVLEEDS, ESL, LSE, UKT2); Romania (UB); Portugal (ICETA, IST); Greece (CERTH, FORTH, ARATOS); Switzerland (EPFL, ETH, UNIGE); Israel (BGU, INPA); Macedonia (HIO); South Africa (CSIR); Austria (EAA); Netherlands (DELTARES, NIOZ); Lithuania (KU); Norway (UiB); Inter. Entity (UNESCO, UNEP); Australia (UNSW); Hungary (SIU); Sweden (UUmea); Venezuela (Provita/IUCN). **And WP** (one or more selection), the possible answers are: ()WP1-Coordination and management; ()WP2-Conceptual Scientific Framework; ()WP3-Earth Observation Data and Processes Infrastructure; ()WP4-Earth Observation Data Generation and Harmonization; ()WP5-In situ Monitoring Data;

()WP6-EO based Ecosystem Modelling; ()WP7-Ecosystem Services; ()WP8-Cross-scale interaction; ()WP9-Requirements of future protected areas; ()WP10- ECOPOTENTIAL Virtual Laboratory Platform; ()WP11-EO supported policy development and integration; ()WP12-Capacity building and knowledge exchange;

- **position relating spatial data life cycle**, the possible answers are: () Data provider; () Data user; () Mainly data provider; () Mainly data user; () Both (equilibrium data provider/supplier and data user/consumer).

As a result of the online questionnaire (see Annex II) **52 responses** were obtained from researchers with a multidisciplinary technical scientific background, mainly Biology/Ecology and Remote Sensing (RS)/Spatial data (SD), as well as Engineering/Technology and Geophysics/Geography. These results indicate a heterogeneous scientific and professional background divided in two major groups: **biologists/ecologists** and **information systems researcher/**professional mainly Remote Sensing/Spatial data modellers and spatial data processing specialists (Fig. 3-2). The ECOPOTENTIAL spatial data users (nearly 70%) have a PhD or Post-PhD degree (Fig. 3-3).

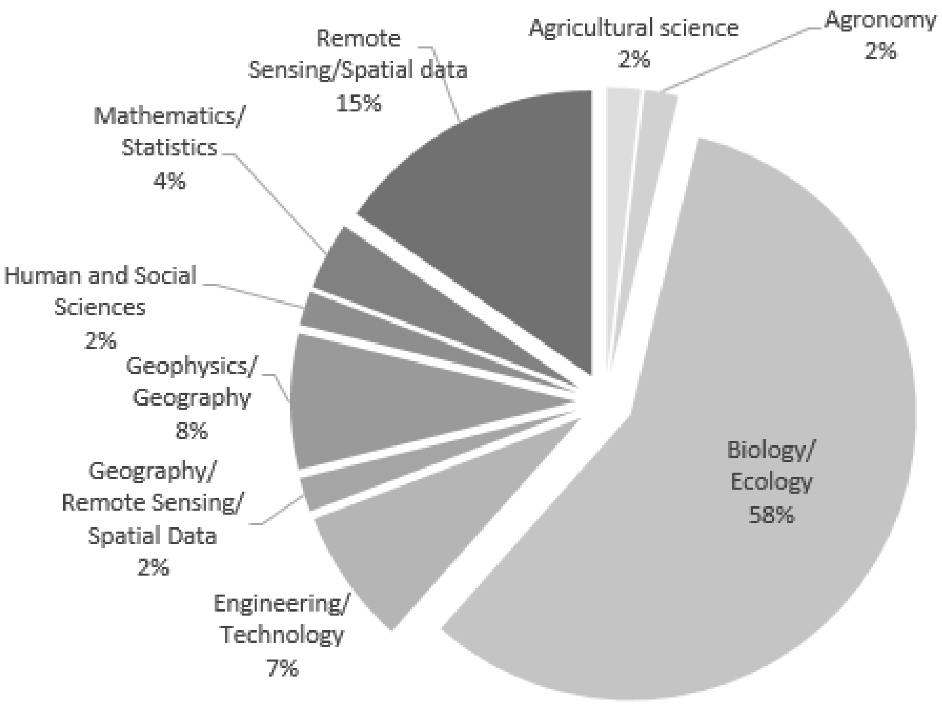


Fig. 3-2. Results for Question 1: Academic background (total 52 responses).

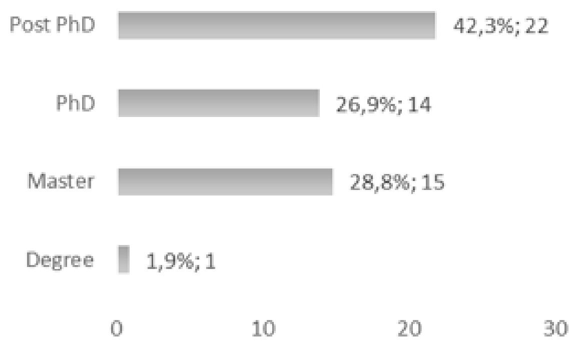


Fig. 3-3. Results for Question 1: Researcher academic degrees (total 52 responses).

The answers were provided by project partners from 15 countries majority Portugal (19%), Spain (19%), Italy (15%) (more response in the countries with mature national SDI) (Fig. 3-4). The questionnaire participants are, mainly involved in “Earth Observation Data Generation and Harmonization” (WP4), “In situ Monitoring Data” (WP5) and “EO based Ecosystem Modelling” (WP6), but also distributed in all ECOPOTENTIAL project WPs (Fig. 3-5).

Regarding the position in the spatial data life cycle, the ECOPOTENTIAL researchers/spatial data users involved in the questionnaire were “mainly data user” or “data user”, with 29% and 19% respectively, or “equilibrium data provider/supplier and data user/consumer” with 31% of responses (Fig. 3-6). The ECOPOTENTIAL project does not includes cadastral/mapping national or regional authorities normally identified as reference spatial data provider. EO/RS specialists consider their contribution as data provider.

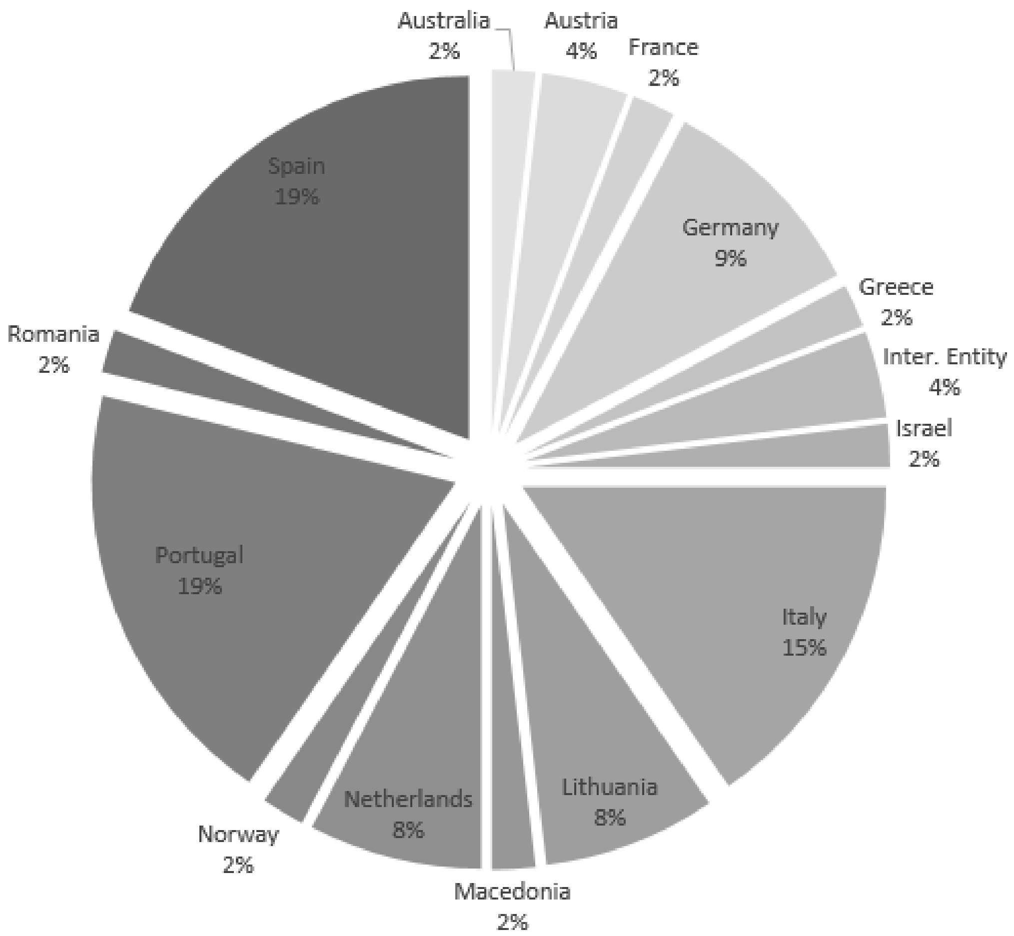


Fig. 3-4. Results for Question 1: Distribution of the involved Researchers by their (working) Country (total 52 responses).

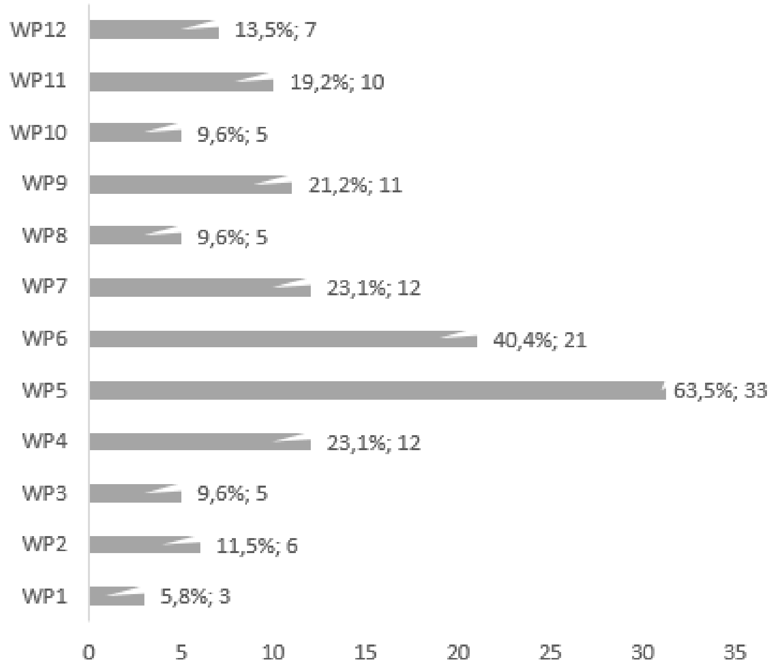


Fig. 3-5. Results for Question 1: Researcher involvement in the WPs (multiple answers possible; total 52 responses).

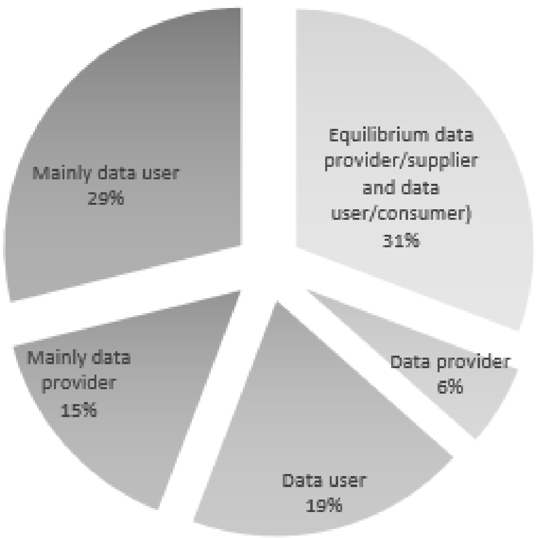


Fig. 3-6. Results for Question 1: Researcher position with respect to spatial data life cycle (total 52 responses).

3.2.2 User's knowledge of spatial (meta)data quality

The purpose of the second group of questions is to understand the user's knowledge about spatial (meta)data quality. The structure of the "yes" or "no" question, combined with the use of a *Likert* scale (for "yes"), reveals opinion/knowledge levels, which can range from 1 to 7 (Fair, Very Low, Low, Medium, Good, Very Good, Excellent).

Question 2.1. Knowledge about ISO standards associated with the spatial data quality. The purpose of this question is to identify the user's knowledge about (spatial) data quality/ domain of data quality concepts (for geographic information or spatial data). The answers reveal generic low knowledge level about ISO standards associated with the spatial data quality: ISO19113:2002; ISO19114:2003; ISO/TS19138; ISO19157:2013 and ISO/TS 19158:2012 (see 37 for details).

The result of online questionnaire indicates that the user's knowledge about spatial (meta)data quality is mostly low for: ISO standards related spatial data quality (Fig. 3-7) and metadata standards/application schemas (Fig. 3-8), although there is a set of user's with good and diversified knowledge. A significant number of ECOP Researchers assumes not to be aware of ISO standards associated with the spatial data quality. At same time, there is a good knowledge about some quality elements (defined by ISO19157:2013) (Fig. 3-9), but in general limited practice of implementation procedures of spatial data quality management along the spatial data life cycle (Fig. 3-10). However, there are user's with good knowledge of data quality, including theory of data quality (guidelines and standards), concepts and data quality elements.

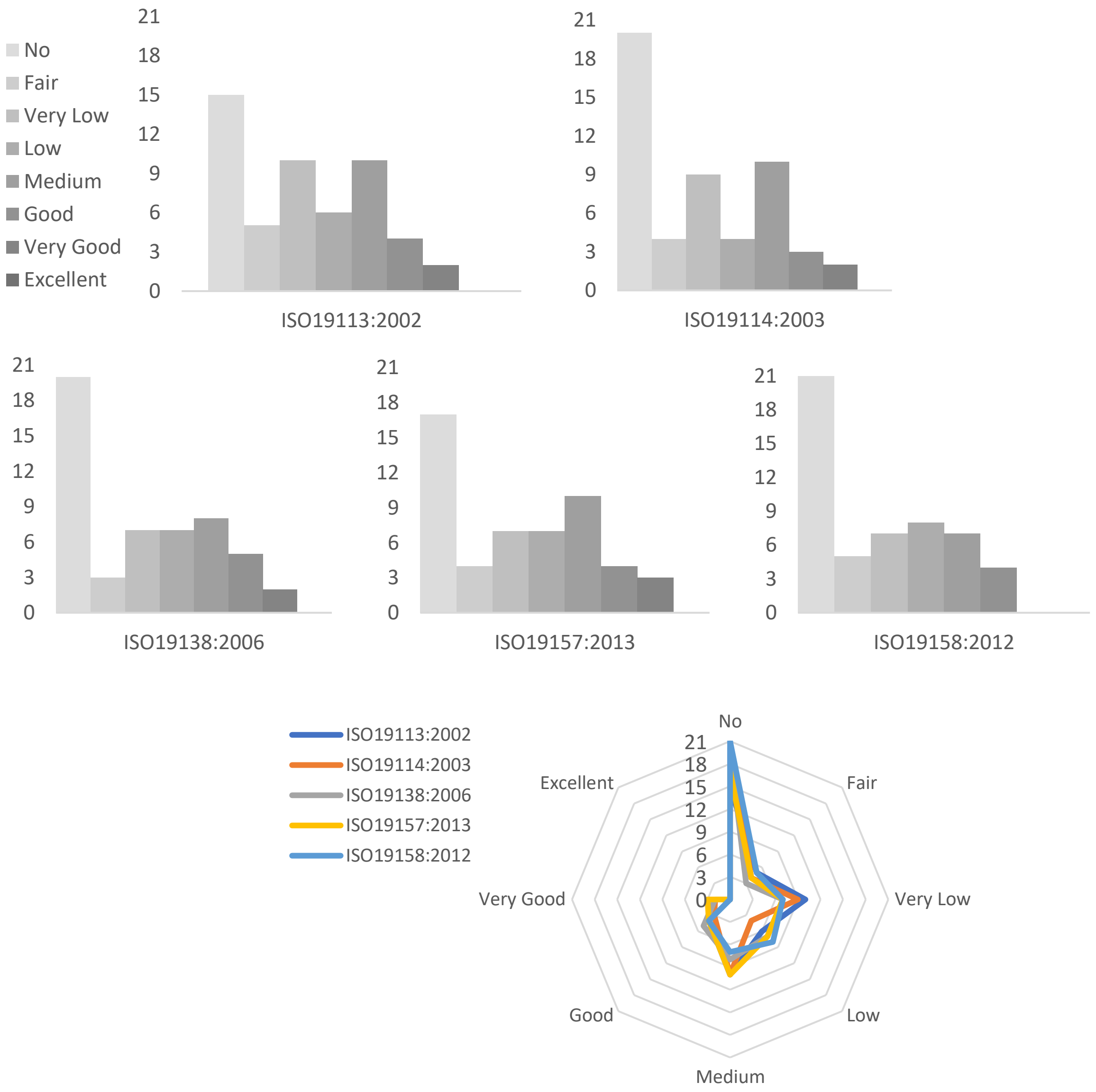
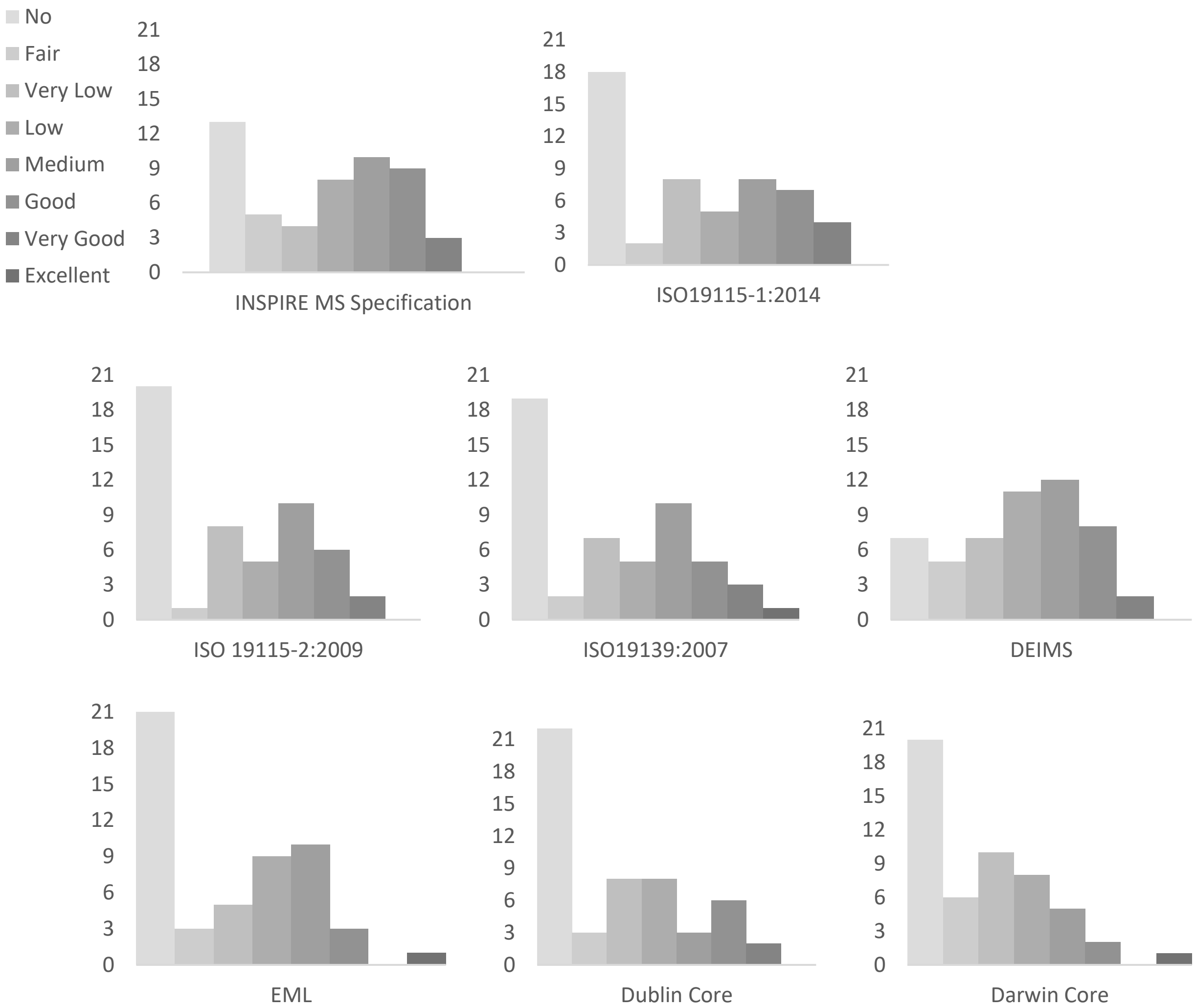


Fig. 3-7. Results for Question 2.1: Knowledge about ISO standards associated with the spatial data quality (total 52 responses).

Question 2.2. Do you know the following metadata standards/application schemas? The purpose of this question is to identify the user's knowledge about spatial data quality assessment and management. The answers reveal knowledge level about metadata standards/application schemas: INSPIRE MS Specifications; ISO 19115-1:2014; ISO 19115-2:2009; ISO/TS 19139:2007; DEIMS; EML; Dublin Core; Darwin Core; ISO 19157:2013 (see 37 for details).



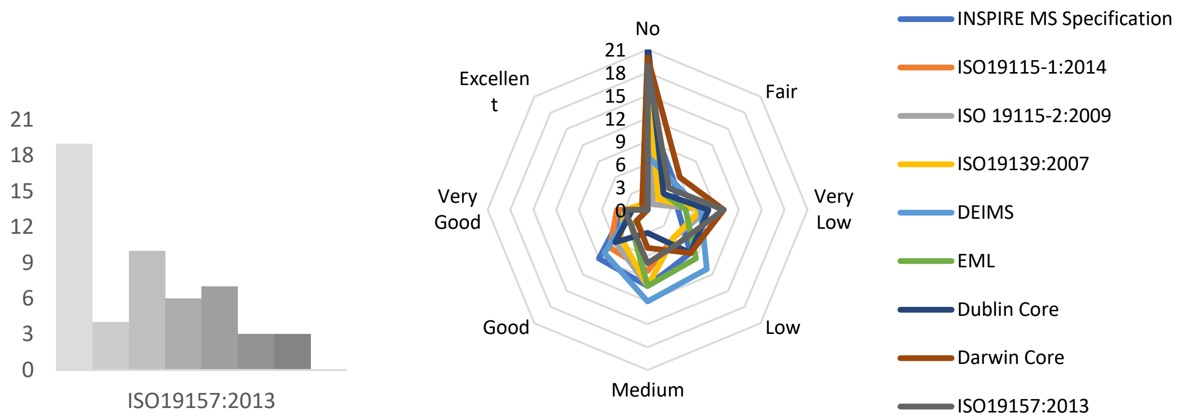
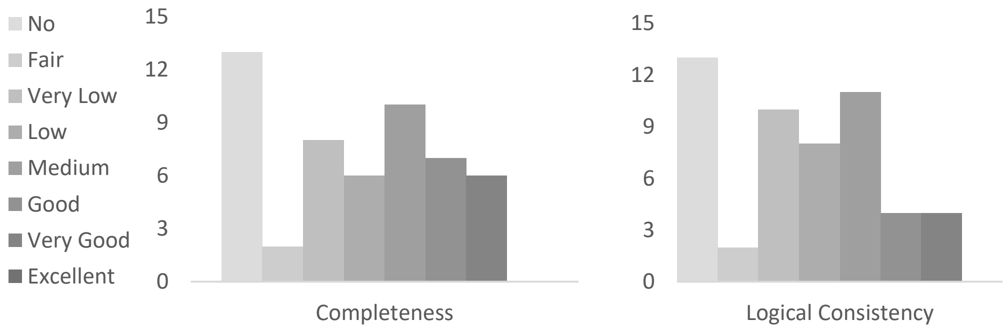


Fig. 3-8. Results for Question 2.2: Knowledge about metadata standards/application schemas (total 52 responses).

The results indicate a significant number of respondents who are not aware of the identified metadata profiles. Among the selected metadata profiles, the majority positive responses refers to the knowledge of spatial metadata profiles and biodiversity data (DEIMS and EML), although only a few (about 1/3) researchers assume good, very good or excellent metadata profile knowledge.

Question 2.3. User knowledge about data quality elements. The purpose of this question is to identify the user's knowledge about data quality element(s) according to ISO19157:2013. The answers (Fig. 3-9) reveal heterogeneous knowledge level about data quality elements: Completeness (presence and absence of features, their attributes and their relationships); Logical Consistency (degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical); Positional Accuracy (accuracy of the position of features); Thematic Accuracy (accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships); Temporal Quality (accuracy of the temporal attributes and temporal relationships of features); Usability Element (degree of adherence of a dataset to a specific set of requirements) (see Tab. 2-1 for definitions).



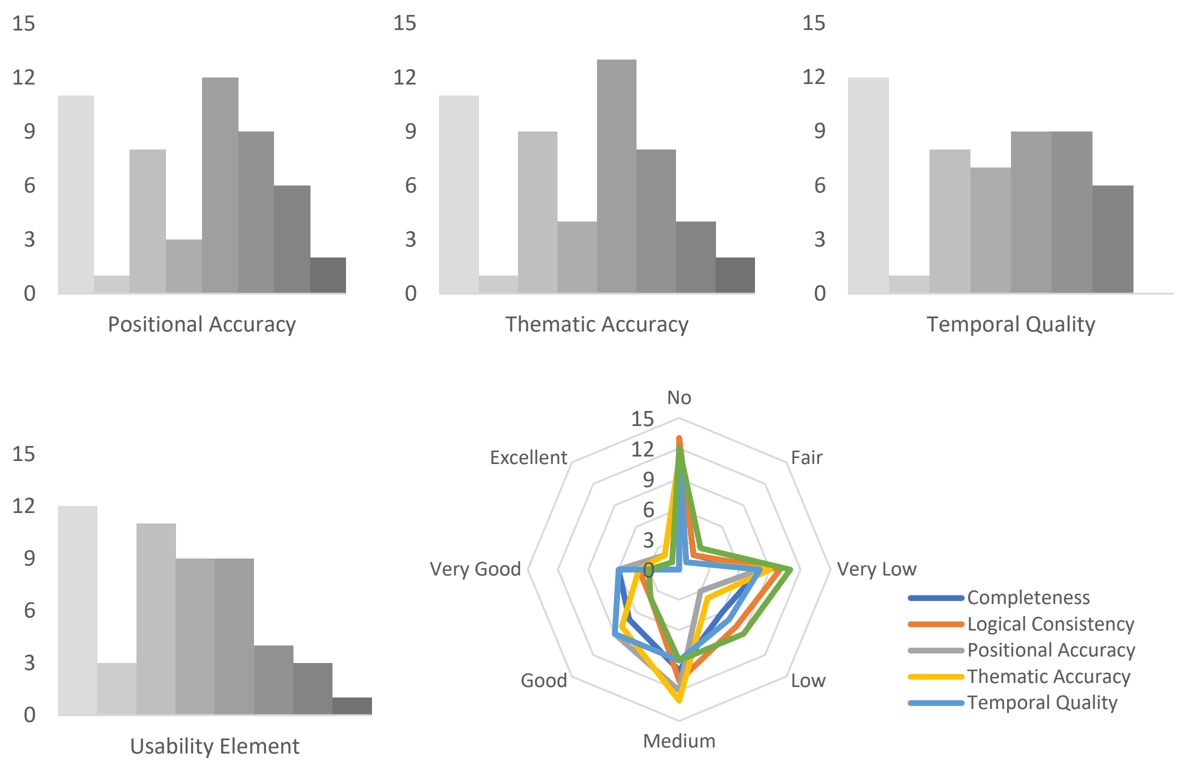


Fig. 3-9. Results for Question 2.3: Knowledge about data quality element(s) (total 52 responses).

The results indicate that more than 12 researchers don´t know data quality elements with emphasis on the temporal quality. Others researchers indicate their knowledge about Completeness, Logical Consistency, Positional, Thematic, Temporal Accuracy and Usability.

Question 2.4. Do you know and/or do you implement procedures of spatial data quality management along spatial data life cycle? The purpose of this question is to identify the user's knowledge about spatial data quality management. The answers reveal the knowledge level about procedures of spatial data quality management along the spatial data life cycle, namely: Data product specification; Data product specification user requirements and provide quantitative quality information; Collecting data (e.g. protocols); Instrumental/technological component (e.g. instrumental selection and calibration); Individual and organizational capacity building (Human component at specification and implementation methods); Data conceptual modelling; Spatial analysis and modelling (e.g. data quality evaluation on Correlative and Process-based models); Data representation, publishing and sharing (e.g. analogic and digital graphical data communication relating data dissemination/accessibility); Evaluating data quality/Reporting data quality (data quality evaluation process/metadata).

Along data the life cycle different users assume low, medium and high knowledge levels regarding the implementation of procedures of spatial data quality management (Fig. 3-10).

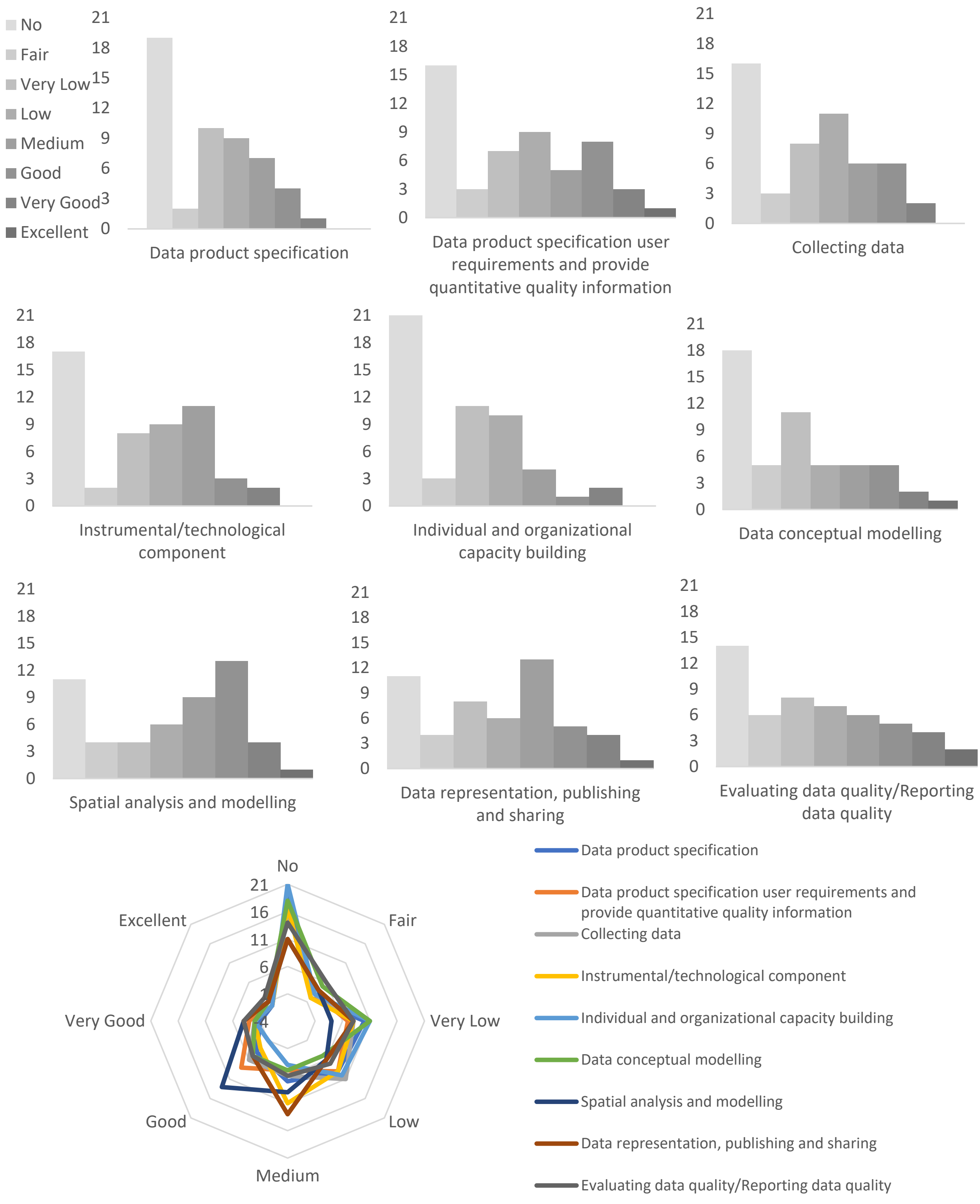


Fig. 3-10. Results for Question 2.4: User's involvement in the implementation procedures of spatial data quality management along spatial data life cycle (total 52 responses).

3.2.3 User's interest towards spatial (meta)data quality

The purpose of the third group of questions is to understand the user's interest on the spatial (meta)data quality.

The structure of the "yes" or "no" question, combined with the use of a *Likert* scale (for "yes"), reveals interest levels, which can range from 1 to 7 (Fair interest, Very Low, Low, Medium, High, Very High, Total interest).

Question 3.1. Do you consider or are you interested in knowing/using spatial data quality elements?

Question 3.2. Do you consider or are you interested in knowing/using spatial data quality assessment methods and tools?

Question 3.3. Do you consider or are you interested in using/participating in spatial data quality management process?

In contrast to the low/medium level of user's knowledge, there is an high interests in knowing/using spatial data quality elements, in knowing/using spatial data quality assessment methods and tools, as well as in using/participating in spatial data quality management process (Fig. 3-11), which may indicate the awareness of the importance of this subject.

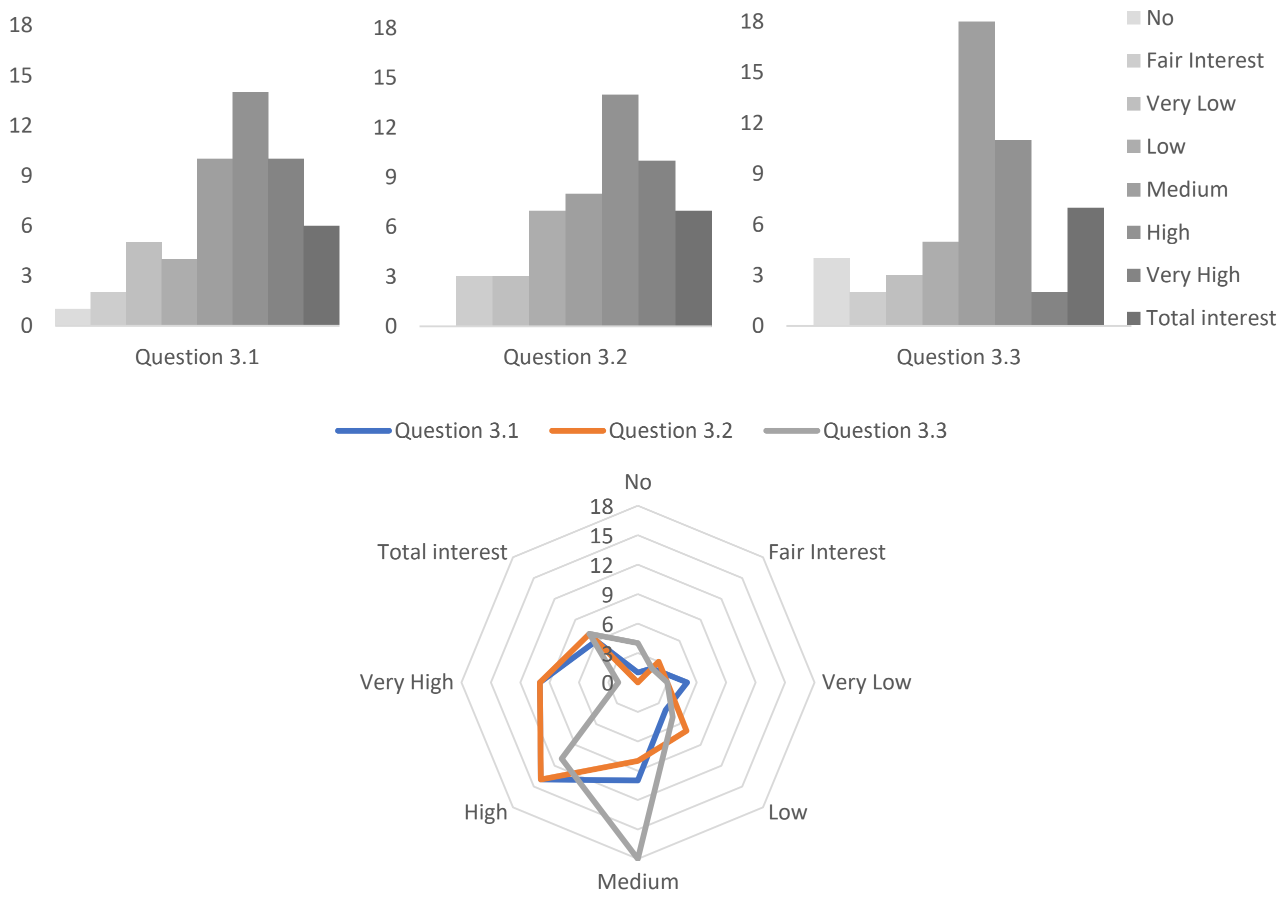


Fig. 3-11. Results for Questions 3.1/3.2/3.3: User's interest of the spatial (meta)data quality (total 52 responses).

3.2.4 User's experience spatial (meta)data quality

The purpose of the fourth group of questions is to understand the user's experience in spatial (meta)data quality.

The structure of the "yes" or "no" question, combined with the use of a *Likert* scale (for "yes"), reveals experience levels, which can range from 1 to 7 (Very Rarely, Rarely, Sometimes, Occasionally, Regularly, Often, Always).

Question 4. On your activities as data provider or data user have you applied data quality evaluation practices/routines? The purpose of this question is to identify the user's experience about practical application of data quality evaluation.

The results of the question 4 showed a high practical experience (92% respondents) about data quality evaluation, which reveals concern about this issue and it is implicit in the workflow (Fig. 3-12).

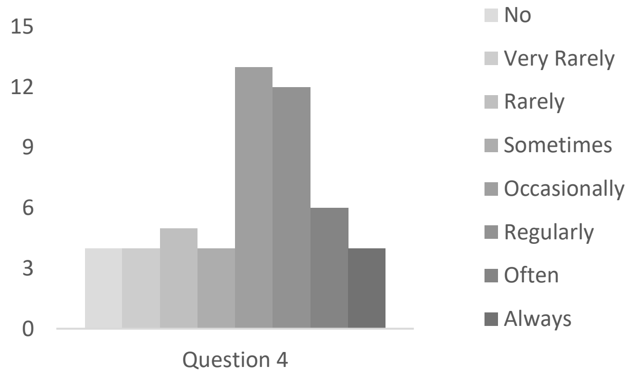


Fig. 3-12. Results for Question 4: User's practical experience about spatial data quality evaluation (total 52 responses).

Question 4.1. Which measures (the type of evaluation) have you applied? The purpose of this question is to identify the user's experience about measures (data quality measures) applied. The answers reveal experience level for: Qualitative measures (e.g. descriptive); Quantitative measures (e.g. number of excess items); Qualitative/ quantitative measures.

Question 4.2. Which evaluation methods (the procedure used to evaluate the measure) have you applied? The purpose of this question is to identify the user's experience about data quality evaluation methods type. The answers reveal experience level for: Direct internal (method of evaluating the quality of a dataset based on inspection of items within the dataset, where all data required is internal to the dataset being evaluated); Direct external (method of evaluating the quality of a dataset based on inspection of items within the dataset, where reference data external to the dataset being evaluated is required); Indirect (method of evaluating the quality of a dataset based on external knowledge); and Aggregation and derivation methods.

Question 4.3. Which results have you obtained in your spatial data quality evaluation process? The purpose of this question is to identify the user's experience about results type obtained. The answers reveal experience level for: Quantitative results; Conformance results; Descriptive results.

The result of the online questionnaire indicates that the user's experience about spatial data quality evaluation is good (Fig. 3-13; Fig. 3-14; Fig. 3-15). More than 50% of users applies "qualitative" (59.6%),

Co-funded by the European Union

"quantitative" (57.7%) or "qualitative/quantitative" (57.7%) measures of data quality evaluation " (Fig. 3-13), with some regularity (including the options: Sometimes, Occasionally, Regularly). We should point out that this issue is away from the expected results. Only 20% users apply measures often or always.

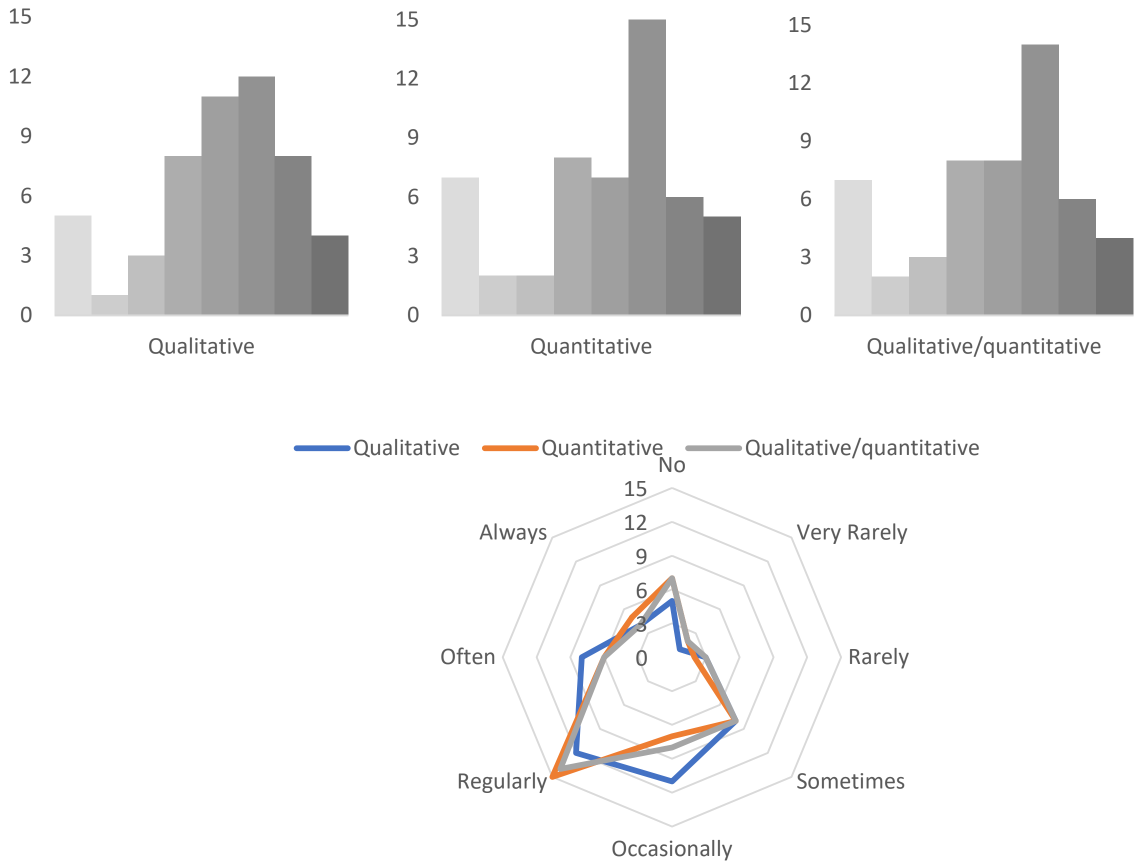


Fig. 3-13. Results for Question 4.1: User's experience about type of measures apply (total 52 responses).

Similarly, regarding the application of direct (internal or external) and indirect methods (Fig. 3-14), it was found that more than 50% of users apply data quality evaluation methods, with some frequency (including the options: Occasionally (71.2%), Regularly (57.7%), Often or Always (50.0%). In relation to results type, quantitative results are more frequently applied/obtained (Fig. 3-15), although conformance and descriptive results are also used.

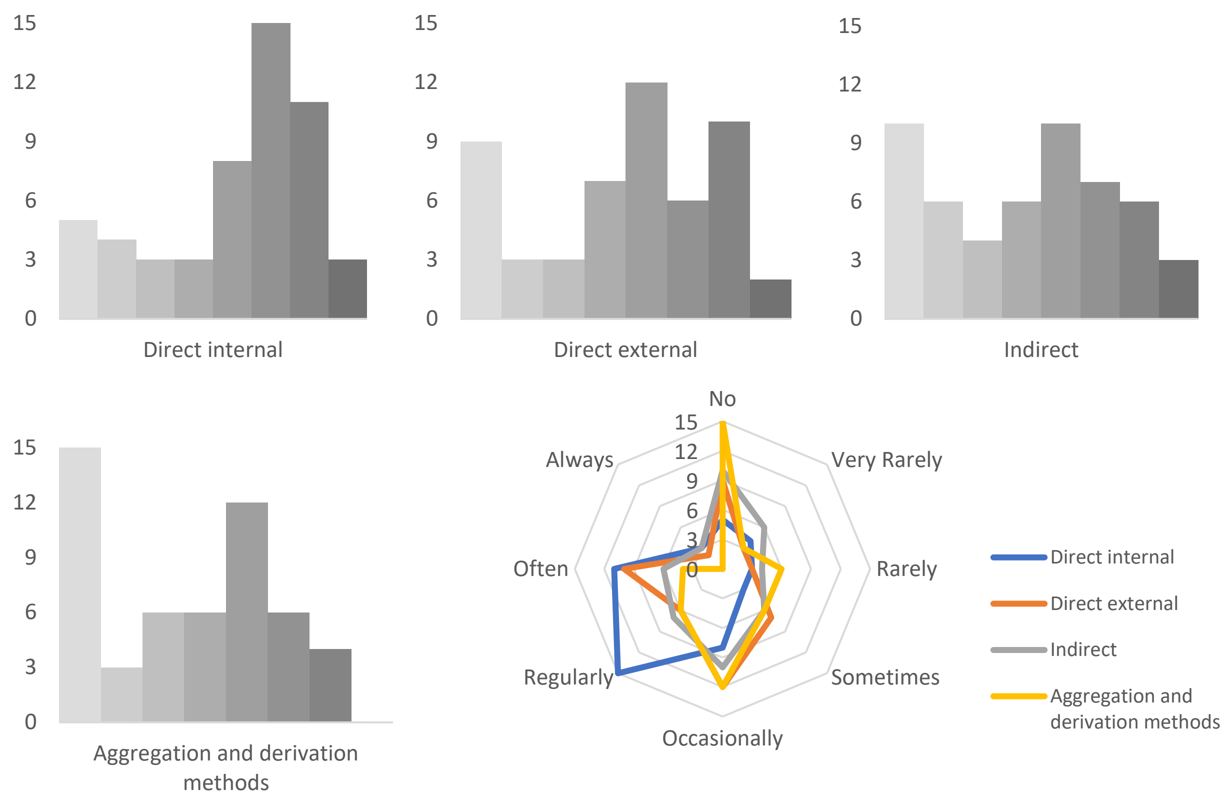


Fig. 3-14. Results for Question 4.2: User's experience about evaluation methods apply (total 52 responses).

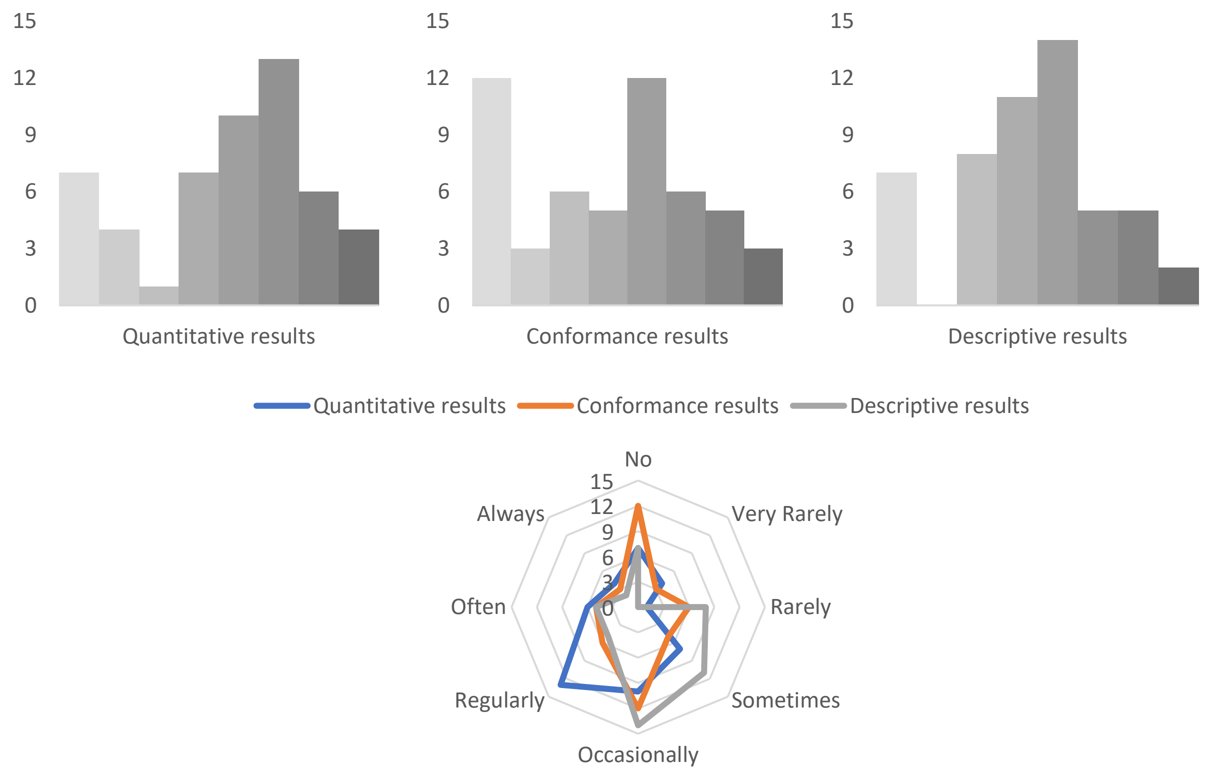


Fig. 3-15. Results for Question 4.3: User's experience about type of results obtain (total 52 responses).

3.2.5 Spatial (meta)data quality user's recognized utility

The purpose of the fifth group of questions is to understand the awareness, interests and willingness to implement data quality routines (utility).

The structure of the "yes" or "no" question, combined with the use of a *Likert* scale (for "yes"), reveal utility levels, which can range from 1 to 7 (Very Rarely, Rarely, Sometimes, Occasionally, Regularly, Often, Always).

Question 5.1. Do you consider or communicate quality assurance (QA)/ quality control (QC):

- Do you **implement procedures** for data QA/ QC?
- Do you **implement and publish documentation about results** of QA/ QC procedures **in metadata**?

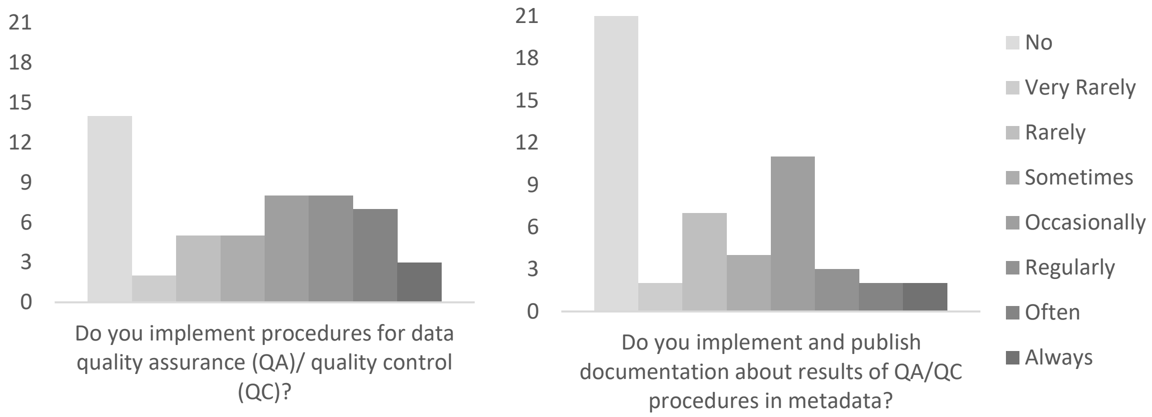


Fig. 3-16. Results for Question 5.1: User's communicate of data quality assurance/control procedures (total 52 responses).

The users show some knowledge and practical experience relating data quality evaluation, but they do not use implementing procedures for data quality assurance (QA)/quality control (QC) (Fig. 3-16).

Question 5.2. Which data quality elements (ISO19157:2013) do you consider decisive:

- **to discover and select input data for applying models and workflows?** The purpose of this question is to identify the data quality elements (ISO19157:2013) that user's consider decisive to discover and select input data for applying models and workflows. The answers reveal utility level for: **Completeness**; **Logical Consistency**; **Positional Accuracy**; **Thematic Accuracy**; **Temporal Quality**; and **Usability Element**, which can range from 1 to 7 (Fair, Very Low, Low, Medium, High, Very High, Decisive).
- **to explore the results of practical/ecological meaning of output data?** The purpose of this question is to identify the data quality elements (ISO19157:2013) that user's consider decisive to explore the results of practical/ecological meaning of output data. The answers reveal utility level for: **Completeness**; **Logical Consistency**; **Positional Accuracy**; **Thematic Accuracy**; **Temporal Quality**; and **Usability Element**, which can range from 1 to 7 (Fair, Very Low, Low, Medium, High, Very High, Decisive).
- **to communicate with end user/technical-political decision makers?** The purpose of this question is

to identify the data quality elements (ISO19157:2013) that user's consider decisive to communicate with end user/technical-political decision makers? The answers reveal utility level for: **Completeness**; **Logical Consistency**; **Positional Accuracy**; **Thematic Accuracy**; **Temporal Quality**; and **Usability Element**, which can range from 1 to 7 (Fair, Very Low, Low, Medium, High, Very High, Decisive).

The results indicate that quality elements are important for finding, analysing/exploring and communicating quality elements throughout the data life cycle, highlighting the temporal quality and usability (Fig. 3-17, Fig. 3-18 and Fig. 3-19).

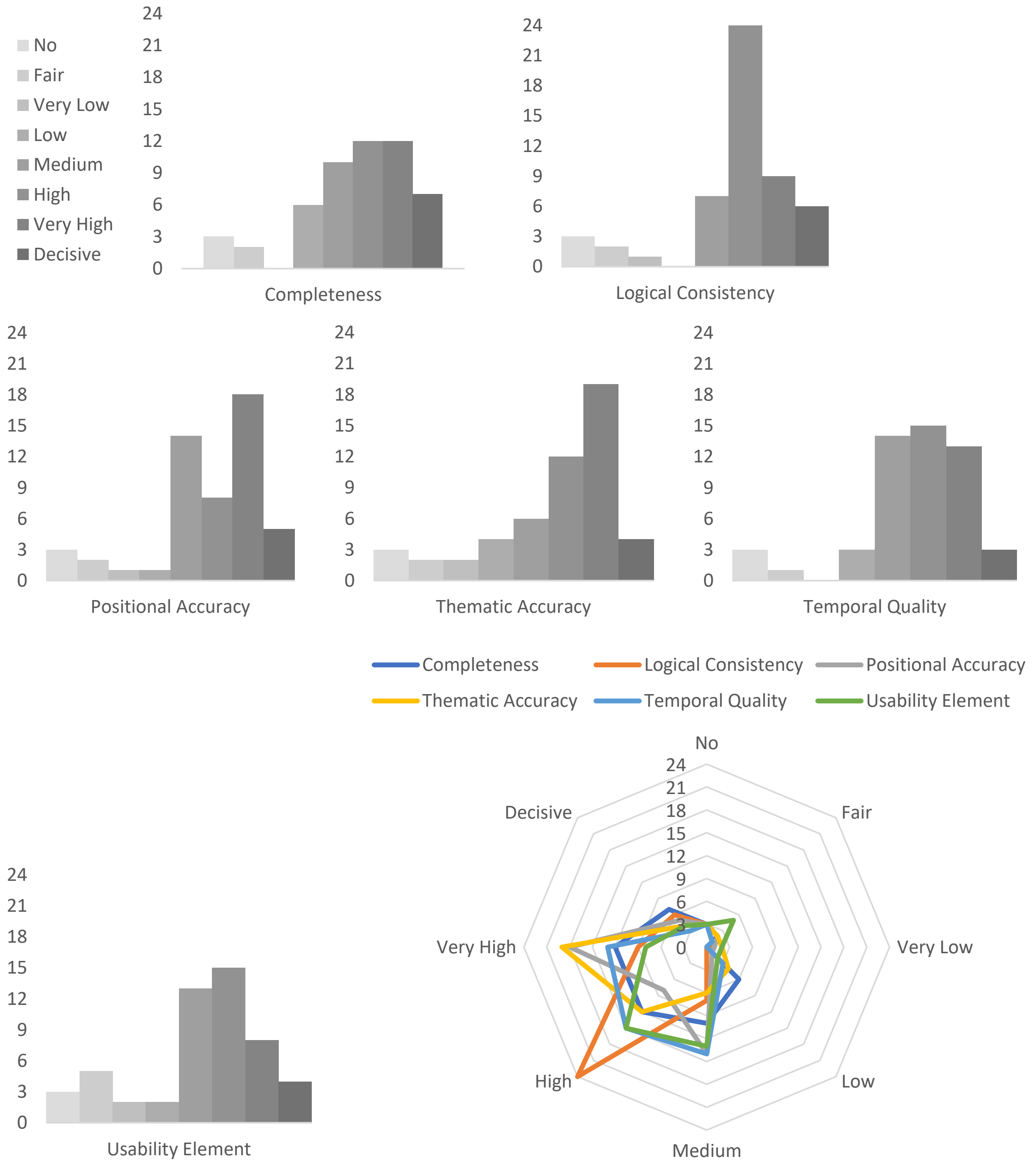


Fig. 3-17. Results for Question 5.2: Data quality elements (ISO19157:2013) that users consider decisive to discover and select input data for applying models and workflows (total 52 responses).

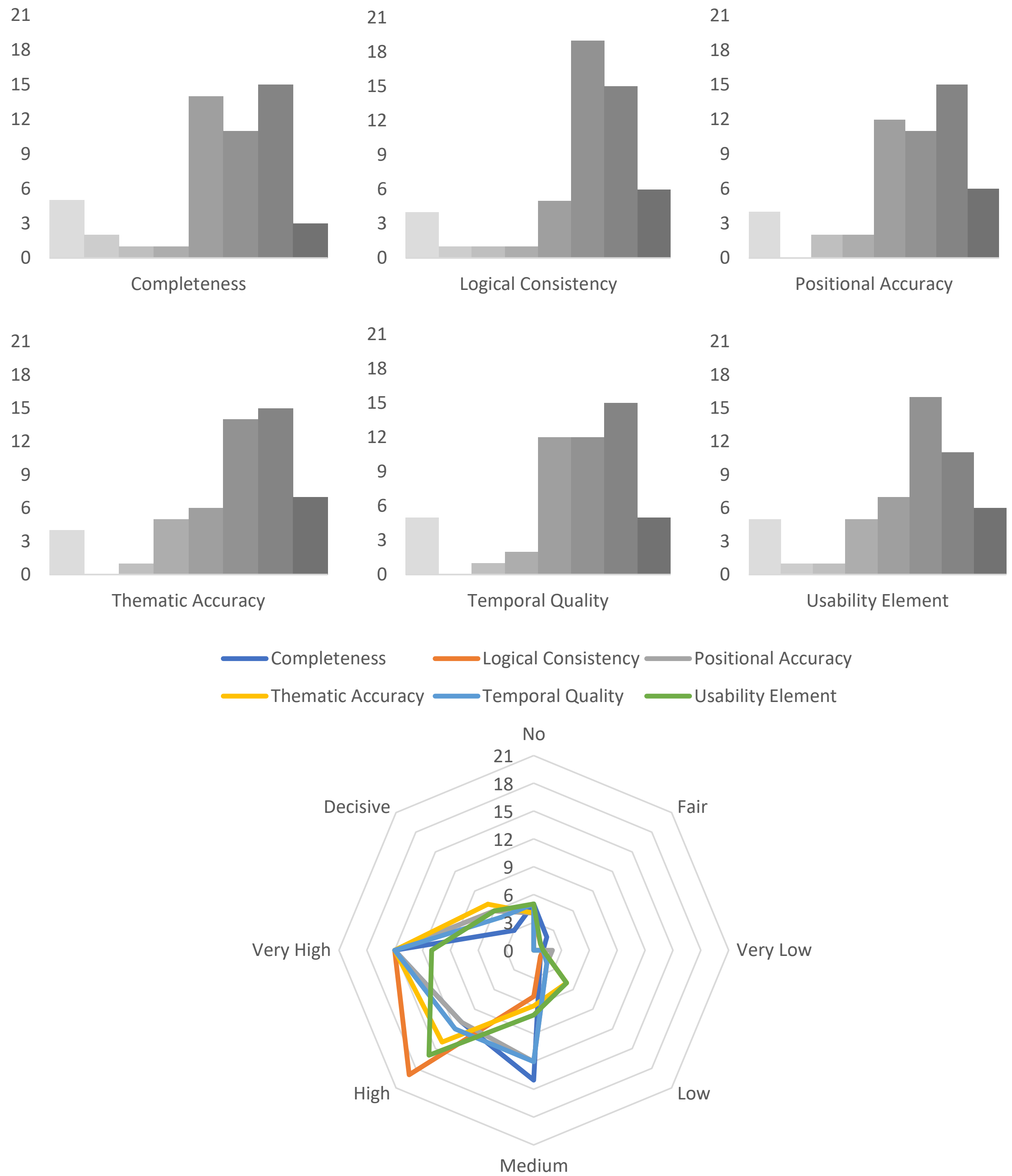


Fig. 3-18. Results for Question 5.2: Data quality elements (ISO19157:2013) that users consider decisive to explore the results of practical/ecological meaning of output data (total 52 responses).

Fig. 3-19. Results for Question 5.2: Data quality elements (ISO19157:2013) user's consider decisive to communicate with end user/technical-political decision makers (total 52 responses).

Question 5.3. Which data quality elements/indicators do you consider important/relevant to incorporate into a metadata profile? *The purpose of this question is to identify the user's interests or recognized quality (ECOPOTENTIAL community):*

- ***Which data quality elements/indicators do you consider important/relevant to incorporate into a metadata profile****: twenty-two suggested answers have been selected. The suggestions are:*

()**Typology** (topic category defining the main data set theme); ()**Taxonomic coverage** (taxonomic classification of the organisms represented in the dataset); ()**Completeness Commission** (number of excess items); ()**Completeness Omission** (number of missing items); ()**Conceptual consistency** (number of items not compliant with the rules of the conceptual schema); ()**Spatial extent** (bounding box defining spatial coverage); ()**Temporal extent** (time interval defining temporal coverage); ()**Lineage** (description of data production methods and overall quality); ()**Methods description** (provides repeated sets of elements that document a series of procedures followed to produce any dataset object); ()**Instrumentation description** (provides information about any instruments used in the data collection or quality control and quality assurance); ()**Sampling description** (provides information about sampling part of the method as measurement frequency, and spatial scale); ()**Quality assurance** (provides information on QA/QC procedures applied for the data); ()**Legal obligation reporting** (provides information whether the dataset has been reported to the local, regional or national bodies to fulfil the obligations from particular legal regulations); ()**Thematic accuracy** (data set thematic accuracy; e.g. number of incorrectly classified features; kappa coefficient); ()**Spatial scale** (equivalent scale or spatial resolution defining the level of detail); ()**Temporal quality** (accuracy of the temporal attributes and temporal relationships of features); ()**Producer recognition** (data producer recognition type); ()**Intellectual Rights** (list of rights management statements for the dataset, or reference a URL (web address) that provides such information); ()**Access and use constraints** (conditions applying to access and use); ()**File format** (distribution file format); ()**Online distribution** (web address is the "navigation section" of a metadata record pointing users to the location (URL) where a dataset can be retrieved directly, or provides information about how to acquire a dataset); ()**Usability** (degree of adherence of a dataset to a specific set of requirements).

- one open answer has been included to receive new suggestions.

The types of metadata elements that the users considered to be the most essential in a spatial metadata profile (Fig. 3-20) were the spatial and temporal extent, the spatial scale, typology, descriptions method, temporal quality and file format. These top choices for types of metadata were closely followed by details on taxonomic coverage and sampling descriptions, but access and use constraints, online distribution, and intellectual property rights were also considered very important. It can be difficult to pinpoint which metadata elements are the most essential to include in a metadata profile, since users do not always find the same information useful, depending on their specific interests.

However, from the questionnaire, it is clear that although all of the suggested 22 elements were considered quite important, few respondents indicated that there were missing elements (3.8%) suggesting 2 more elements: Spatial Reference System; keywords/keywords vocabulary. This result is promising as it indicates that we have identified the elements most essential to users within the project. However, it is necessary to extend this type of study to other user groups, so that this process can be further standardized and suited to a wider variety of scientists.

The responses of question 5.3, indicate which spatial data elements/indicators are considered important, critical and relevant to incorporate into a metadata profile in order to support/facilitates data quality assessment and management in scientific collaborative network management (Fig. 3-20).

ECOPOTENTIAL community presents awareness and need to document their data in metadata as a way to allow data identification (users), data dissemination (producers) and data sharing (collaborative networks, data users/providers).

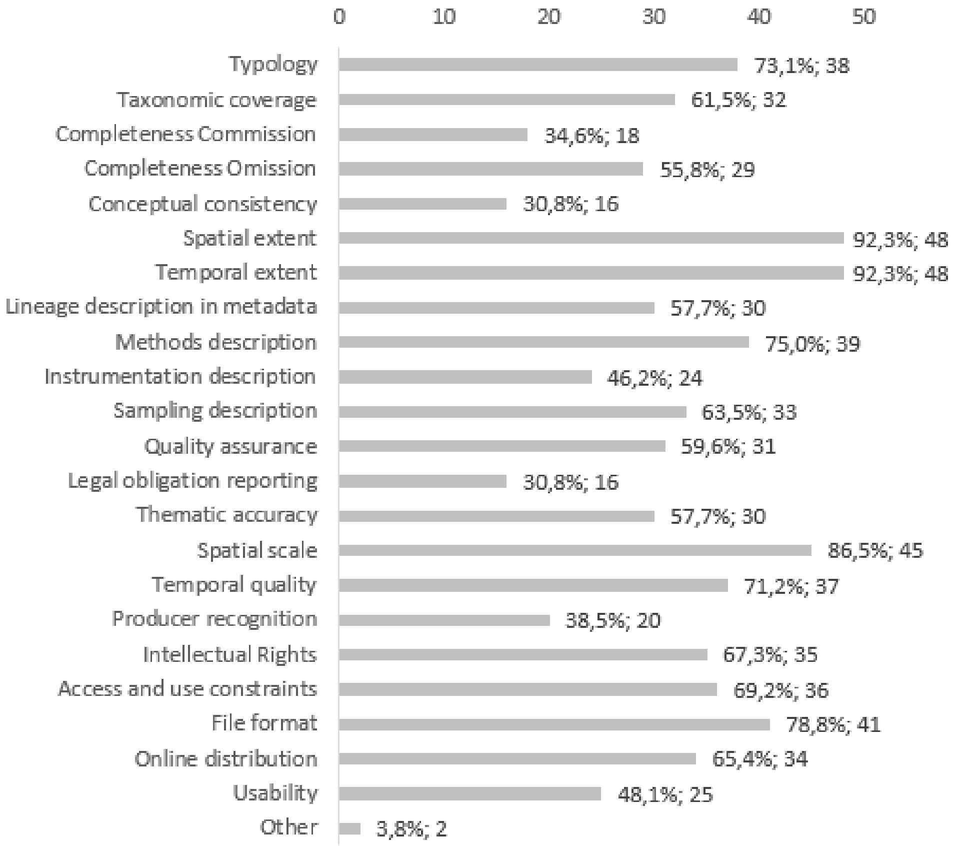


Fig. 3-20. Results for Question 5.3: Relevant spatial elements/indicators to incorporate into a spatial metadata profile in order to support/facilitates data quality assessment and management in scientific collaborative network management (total 52 responses).

These results indicates interests specific needs, experiences and recognized quality utility and define users groups (e.g., modellers, GIS analysts, remote-sensing specialists, in situ data collection experts, …). Each group presents differentiated domains in quality evaluation and management with practical implication on adequate (meta)data models and metadata profile selection as well as, in data quality assessment methods and tools analysis specification. The synthesis of this line questionnaire and user´s group profiles show different languages between technological and ecologists ECOPOTENTIAL researcher and main and transversal interests. The end user should be responsible to define/select a set of quality element descriptors (e.g., resolution, accuracy, spatial and/or temporal consistency) for typified application contexts common to their specific technical-scientific application domain and also define, predefined quality expected values. The on-line questionnaire, meetings, practical exercises and other contacts with ECOPOTENTIAL researchers indicate the difficulties in define relevant quality elements and mainly, expected quality indicators in datasets used in socio-ecological modelling as their associated results.

The questionnaire results relevant for activities in WP5, namely when: (1) defining the fields/attributes of quality elements to include in the selected metadata profile (Task 5.3/5.5); (2) specifying and developing methods (external evaluation) and tools (routines and information/technological application) of spatial data quality assessment (Task 5.5); and (3) devising proposals for implementation of quality management processes (WP5 and recommendations for others WPs/ECOPOTENTIAL Project).

3.3 Metadata quality elements and (meta)data management

This alternative and complementary metadata profile assessment (made in cooperation with Task 5.3) contributes to the discussion of new fields (quality elements) to be included in the selected metadata profile in order to effectively allow/perform fitness for use spatial data quality assessment and data quality management supported on (in)direct evaluation methods (ISO 19157: 2013).

3.3.1 Proposal of relevant metadata fields related to spatial data quality evaluation

The scientific and technological community of (spatial) data quality assessment presents novel conceptual approaches and methodological frameworks, namely in the meta-evaluation of external and (in)direct quality by the end-user(s), taking advantage of documentation and communication possibilities needs expanded quality elements in metadata profiles.

The need to standardize languages, define approaches and concepts and procedures is established by ISO 19157 in its relation to the control and quality assurance of data supply (ISO 19158; Geographic information - Quality assurance of data supply). The development of metadata profiles and metadata catalogues according to global (ISO 19115 and ISO 19139), international (INSPIRE) or thematic (DEIMS) standards improves (in)direct quality assessment processes and promotes data quality practices and routines. The user perspective or fitness for use quality evaluation motivates the development and implementation of user oriented quality evaluation routines for a rapid/agile and adequate assessment external quality of pre-existing data based on (spatial) metadata (Task 5.3) quality evaluation.

A comparative analysis of different metadata profiles indicates insufficiency and omissions relating of data quality aware elements fields (Tab. 3-1). INSPIRE MD profile is more demanding. The metadata comparative analysis indicates the importance of increasing and detailing the quality elements fields related to descriptors of quality elements according to ISO19157:2013, in order to improve and guarantee quality control (and report by metadata) in the production and maintenance (scope and frequency of updating) of quality throughout the data life cycle (data management) (Tab. 3-2). The analysis (Tab. 3-1 and Tab. 3-2) indicates the importance and possibility of suggesting/adding new data quality (sub)elements (recommended by ISO19157) into the metadata profiles in order to complement the missing elements as well as to value/support the data quality evaluation exercises/processes. We recommend the addition of new fields (quality elements) to be included in the DEIMS-SDR MD [Data quality elements], or the adaptation of some profile fields, namely its structure and content [17. Dataset methods; 18. Dataset Instrumentation description; and 19. Dataset Sampling description], and whether it is mandatory or optional [10. Dataset Access and Use constraints; Dataset Intellectual Rights, and 12. Dataset Online distribution], in order to improve the possibilities for assessing the quality of spatial data.

ECOPOTENTIAL community presents awareness and need to document their data in metadata as a way to allow data identification (users), data dissemination (producers) and data sharing (collaborative networks, data users/providers).

The typology of attributes (e.g. value or free text), the number of elements, the correct completion of the metadata fields (according to standards) define and limit the number of quality indicators that we can infer about the metadata. An example is the case of element "lineage" related to quality characteristics in the first versions of ISO19115 and INSPIRE guidelines (in which INSPIRE is more demanding because it is mandatory), whose attribute is free text. Although it is possible to search in this text (e.g. in developed ThemisE platform) for keywords or regular expressions, a user with a lack of knowledge of the content type of this field will not be able to perform an adjusted query/request. This exercise is even more complex in the case of the DEIMS-SDR MD model, where several metadata elements (i.e. "Dataset Methods description", "Dataset Instrumentation", "Dataset Sampling description" and "Quality assurance") can be integrated into the field "Lineage". This attribute in this profile "describes the history of a dataset and recounts the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form; in general, non-quantitative information is illustrative for users and can help assessing the quality of a dataset, especially in cases where it is used for a particular application that differs from the intended application ".

The spatial data quality scope, elements and indicators definition implies answers to the following questions (Fig. 3-21): (i) How to select/define the relevant metadata quality indicators (e.g., resolution, accuracy, spatial and/or temporal consistency)? What data quality elements selected from quality elements/components are included in ISO 19157?; (ii) What are the indicators which the data providers can fulfil?; (iii) Is it necessary to consider distinctive quality indicators for different datasets types?; (iv) How to select an appropriate metadata profile for ECOPOTENTIAL Project incorporating relevant data quality descriptors?. The advances in metadata profile have implication on technological requirements and development, in terms of technological architecture and functional capabilities, but also in terms of user interface design where the user can select quality indicator, define expected values and visualize report results. Will it be advantageous to define user groups (e.g., modellers, GIS analysts, remote-sensing specialists, in situ data collection experts) where each group should be responsible to define/select a set of quality element descriptors (e.g., resolution, accuracy, spatial and/or temporal consistency) for typified application contexts common to their domain? Will it be advantageous to define predefined values?. These questions permit supporting use and user requirements definition according to the ECOPOTENTIAL data quality evaluation platform specifications (see section 4).

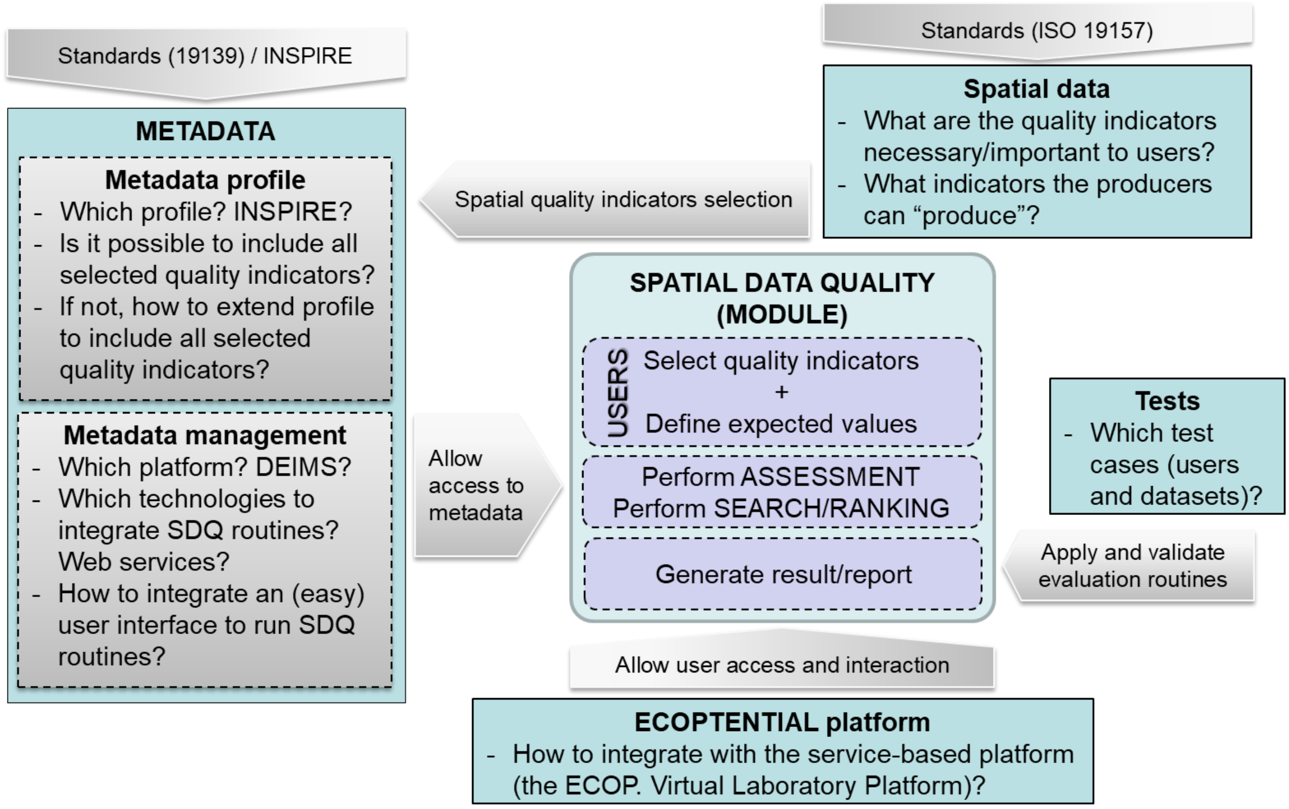


Fig. 3-21. Metadata profile and metadata management related to spatial data quality evaluation.

The identification, analysis and specification of requirements for the user-oriented quality evaluation routines, supported on metadata, provides: (1) analyse candidate metadata profiles and documentation (Tab. 3-1 and Tab. 3-2); (2) evaluate the current DEIMS-SDR MD model in terms of adequacy for data quality assessment (Tab. 3-3); (3) inquire the ECOPOTENTIAL community (Annex I and section 3.2) of data providers and data user's (e.g. WP4, WP5, WP6, WP9, WP10, …) about their knowledge, their practical experience and their awareness (utility) of data quality evaluation routines (results, in particular fifth group of questions) (Annex II and section 3.2.5); and (4) provides the proposal of relevant fields/elements related to data quality evaluation routines (Tab. 3-4).

The present proposal combines the selection of the metadata elements that best fit the metadata data quality assessment (i.e. the application of an indirect evaluation method, according to ISO19157 ...), and which allows the external quality evaluation (users) in data adequacy perspective to different contexts of application/use. The proposal data quality elements (Tab. 3-4) should support, consider and result in data quality evaluation procedure, provide additional information about the scope (in this case: **spatial datasets and datasets series, not services**), frequency of updating (maintenance information), lineage (process step, …), results of metadata conformity[32], information about Data quality evaluation process and Data quality Element(s) evaluated. The metadata/quality relevant fields include descriptive, quantitative and qualitative elements along main characteristics quality.

Tab. 3-4 Proposed metadata elements for spatial data quality evaluation procedure (task 5.5, also reported for D5.2,

[32] The metadata shall, in conformance to [INSPIRE Directive], include a statement on the degree of conformity with the specifications against which its conformity has been evaluated.

updated).

Main Characteristics Quality (see 45 for definitions)	**Metadata/Quality Element**	**Metadata element**	**Definition**
Definition	INSPIRE Spatial Data themes	Keywords	Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject (in this case, INSPIRE Spatial Data themes), or other keywords related the controlled vocabulary implemented by EnvThes-thesaurus for long term ecological research (LTER), monitoring, experiments EnvThes and other environmentally related thesauri as EUNIS Habitats and others.
	Scope	Resource type (scope)	Scope to which metadata applies "MD_ScopeCode" (Data type – see annex B.5.25 of ISO 19115)
	Typology	Topic category	The topic category is a high-level classification scheme to assist in the grouping and topic-based search of available spatial data resources "MD_TopicCategory", according to the ISO 19115.
		Taxonomic coverage	Taxonomic Biological classification of the organisms represented in the dataset or dataset series.
		Date of publication	This is the date of publication of the resource when available, or the date of entry into force.
		Maintenance information	This field refers to the information about the scope and frequency of updating.
Data quality	Completeness	Commission	Excess data present in a dataset or dataset series.
		Omission	Data absent from a dataset or dataset series.
	Logical Consistency	Conceptual consistency	Adherence to rules of the conceptual schema.
		Domain consistency	Adherence of values to the value domains.
		Format consistency	Degree to which data is stored in accordance with the physical structure of the dataset or dataset series.
		Topological consistency	Correctness of the explicitly encoded topological characteristics of a dataset.
Coverage	Spatial/Geographic	Geographic bounding box	This field refers to the geographical scope of the dataset, particularly whether the dataset covers all or just a portion of the study area. 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.
	Temporal	Temporal extent	The temporal extent defines the time period covered by the content of the resource.
	Geographic (Research site)	Geographic (Research site)	The Geographic reference for the dataset is done by the entity type Research site, which is the location, where specific observations are done.
Lineage	Lineage (process step, …)	Lineage	Lineage describes the history of a dataset and recount the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form.
		Methods description	Provides repeated sets of elements that document a series of procedures followed to produce any dataset object.
		Instrumentation description	Provides information about any instruments used in the data collection or quality control and quality assurance.
		Sampling description	Provides information about sampling part of the method as measurement frequency.
		Quality assurance	Provides information on QA/QC procedures applied for the data.
		Legal obligation reporting	Provides information whether the dataset has been reported to the local, regional or national bodies to fulfil the obligations from particular legal regulations.

Main Characteristics Quality (see 45 for definitions)	Metadata/Quality Element	Metadata element	Definition
Data quality	Positional Accuracy	Absolute external positional accuracy	Closeness of reported coordinate values to values accepted as or being true.
		Relative internal positional accuracy	Closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true.
		Gridded data positional accuracy	Closeness of gridded data spatial position values to values accepted as or being true.
	Temporal Quality	Accuracy of a time measurement	Closeness of reported time measurements to values accepted as or known to be true.
		Temporal consistency	Correctness of the order of events.
		Temporal validity	Validity of data with respect to time.
	Thematic Accuracy	Thematic classification correctness	Comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data).
		Non-quantitative attribute correctness	Measure of whether a non-quantitative attribute is correct or incorrect.
		Quantitative attribute accuracy	Closeness of the value of a quantitative attribute to a value accepted as or known to be true.
Precision		Spatial resolution	Spatial resolution refers to the level of spatial detail of the dataset. It shall be expressed as a set from zero to many resolution distances (typically for gridded data and imagery-derived products) or equivalent scales (typically for maps or map-derived products). An equivalent scale is generally expressed as an integer value expressing the scale denominator. A resolution distance shall be expressed as a numerical value associated with a unit of length.
		Spatial Reference System	This field refers to the geographical reference system of the dataset.
Legitimacy		Author	This field refers to the institution or individual that produced the dataset (Responsible organisation).
		Intellectual Rights	List of rights management statements for the dataset.
Accessibility		Property	This field refers to the property of the dataset being necessary to state if there are any conditions applying to its access and use.
		File format	Type of file of the dataset or dataset series (distribution format).
		Resource locator	Location (address) for on-line access using a Uniform Resource Locator address or similar addressing scheme.
Data quality	Usability	Usability element	Degree of adherence of a dataset to a specific set of requirements.

Table 3-4 (cont.)

INSPIRE Directive is based on ISO19115 and ISO19139 standard(s), and aims to make available relevant, harmonized and quality geographic information to support the formulation, implementation, monitoring and evaluation of policies and activities, which have a direct or indirect impact on the environment. The 'Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS19139:2007' (Date of publication: 2017-03-02) list the requirements and rules to the implementation of metadata for **spatial datasets or datasets series** and services**,** for all spatial data themes listed in the Directive 2007/2/EC Annex I, II or III (Tab. 3-6), included in Regulations 1205/2008 and 1089/2010.

The Commission Regulation (EC) No 1205/2008 of 3 December 2008, implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata defines '*The value domain of each metadata element is necessary to ensure interoperability of metadata in a multilingual context and that value domain should be able to take the form of free text, dates, codes derived from international standards, such as language codes, keywords derived from controlled lists or thesauri, or character strings*' (EC, 2008).

The Commission Regulation (EU) No 1089/2010 of 23 November 2010, implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial datasets and services defines, '*In order to achieve interoperability and benefit from the users' and producer's communities, when appropriate, international standards are integrated into the concepts and definitions of the elements of spatial data themes listed in the Directive 2007/2/EC Annex I, II or* III (EU, 2010).

This Technical Guidance (for ISO/TS19139 based XML format implementation) defining metadata in compliance with the INSPIRE implementing Rules (EC, 2017). This Technical Specification provides Extensible Markup Language (XML) schemas, namely: (i) Geographic Common extensible markup language (GCO), the XML Schema Implementation of ISO 19103 concepts; (ii) Geographic Metadata extensible markup language (GMD), the XML Schema Implementation of ISO 19115 concepts; (iii) Geographic Spatial Schema extensible markup language (GSS), the XML Schema Implementation of ISO 19107 concepts; (iv) Geographic Spatial Referencing extensible markup language (GSR), the XML Schema Implementation of ISO 19111 concepts; (v) Geographic Temporal Schema extensible markup language (GTS), the XML Schema Implementation of ISO 19111 concepts; (vi) Geographic Metadata XML Schema (GMX), the XML Schema Implementation of the concepts defined in ISO/TS 19139:2007. The namespaces of these XML Schema Implementations is http://www.isotc211.org/2005/ followed by the corresponding lowercase acronym (i.e. prefix): http://www.isotc211.org/2005/gmd for GMD.

The required INSPIRE metadata elements for spatial datasets and datasets series (task 5.5 does not include services, INSPIRE Implementing Rules: Regulation 1205/2008 or Regulation 1089/2010, and INSPIRE metadata element obligation type: mandatory or conditional; for metadata conformity), as well as, ISO/TS 19139 path (namespace[33] or package[34]) based on geographic metadata extensible markup language *.gmd) and domain value(s), are mapped in the following table (Tab. 3-5).

Tab. 3-5. Overview of the required INSPIRE Metadata elements (mandatory or conditional) for Spatial **datasets** and **datasets series**, and INSPIRE Implementing Rules for metadata (ISO/TS19139:2007 adapted).

Implementation rule	**Metadata Element Name**	**INSPIRE obligation/ condition/ note**	**ISO/TS 19139 path**	**Domain**
[Regulation 1205/2008]	Resource title	Mandatory	identificationInfo[1]/*/citation/*/title # Data type (and ISO19115 no.) CharacterString	Free text
[Regulation 1205/2008]	Resource abstract	Mandatory	identificationInfo[1]/*/abstract # Data type (and ISO19115 no.) CharacterString	Free text
[Regulation 1205/2008]	Resource type	Mandatory	hierarchyLevel # Data type (and ISO19115 no.) MD_ScopeCode	CodeList (see annex B.5.25 of ISO 19115)

[33] Namespace is a collection of names, identified by a URI reference, that are used in XML documents as element names and attribute names.
[34] Package is a general purpose mechanism for organizing elements into groups (e.g. Metadata entity set information, Constraint information, …).

Implementation rule	Metadata Element Name	INSPIRE obligation/ condition/ note	ISO/TS 19139 path	Domain
[Regulation 1205/2008]	Resource locator	Mandatory if a URL is available to obtain more information on the resources and/or access related services	distributionInfo/*/transferOptions/*/onLine/*/linkage # Data type (and ISO19115 no.) URL	URL (IETF RFC1738 and IETF RFC 2056)
[Regulation 1205/2008]	Unique resource identifier	Mandatory	identificationInfo[1]/*/citation/*/identifier # Data type (and ISO19115 no.) 205. MD_Identifier	URI (IETF RFC 3986)
[Regulation 1205/2008]	Resource language	Mandatory if the resource includes textual information	identificationInfo[1]/*/language # Data type (and ISO19115 no.) LanguageCode (ISO/TS 19139)	Codelist (See ISO/TS 19139) based on alpha-3 codes of ISO 639-2. Use only three-letter codes from in ISO 639-2/B (bibliographic codes), The list of codes for the 24 official EU languages is: Bulgarian – bul Irish – gle Croatian – hrv Italian – ita Czech – cze Latvian – lav Danish – dan Lithuanian – lit Dutch – dut Maltese – mlt English – eng Polish – pol Estonian – est Portuguese – por Finnish – fin Romanian – rum French – fre Slovak – slo German – ger Slovenian – slv Greek – gre Spanish – spa Hungarian – hun Swedish – swe The list of all the codes is defined at http://www.loc.gov/standards/iso639-2/ Regional languages also are included in this list.
[Regulation 1205/2008]	Topic category	Mandatory	identificationInfo[1]/*/topicCategory # Data type (and ISO19115 no.) MD_TopicCategory	Enumeration (See B.5.27 of ISO 19115 or Part D 2 of [Regulation 1205/2008])
[Regulation 1205/2008]	Keyword value	Mandatory	identificationInfo[1]/*/descriptiveKeywords/*/keyword # Data type (and ISO19115 no.) CharacterString	Free text
[Regulation 1205/2008]	Originating controlled vocabulary	Conditional: Mandatory for each keyword if the keyword value originates from a controlled vocabulary	identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName # Data type (and ISO19115 no.) identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName	The following properties are expected: -> Title (characterString and free text) -> Reference date (CI_Date): dateType: creation, publication or revision date: an effective date
[Regulation 1205/2008]	Geographic bounding box	Mandatory	identificationInfo[1]/*/extent/*/geographicElement/*/westBoundLongitude identificationInfo[1]/*/extent/*/geographicElement/*/eastBoundLongitude identificationInfo[1]/*/extent/*/geographicElement/*/southBoundLatitude identificationInfo[1]/*/extent/*/geographicElement/*/northBoundLatitude #	-180.00 ≤ westBoundLongitude ≤ 180.00 -180.00 ≤ eastBoundLongitude ≤ 180.00 -90.00 ≤ southBoundingLatitude ≤ 90.00 -90.00 ≤ northBoundingLatitude ≤ 90.00

Implementation rule	Metadata Element Name	INSPIRE obligation/ condition/ note	ISO/TS 19139 path	Domain
			Data type (and ISO19115 no.) Decimal	
[Regulation 1205/2008]	Temporal reference	At least one of Temporal extent, Date of publication, Date of last revision or Date of creation must be given		
[Regulation 1205/2008]	Temporal extent	Conditional: At least one temporal reference is required	identificationInfo[1]/*/extent/*/temporalElement/*/extent # Data type (and ISO19115 no.) TM_Primitive	As described in ISO 19108
[Regulation 1205/2008]	Date of publication	Conditional: at least one date of publication / date of creation / date of revision is required	identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date # Data type (and ISO19115 no.) 393. CI_Date	Described in ISO 19108 and ISO 8601
[Regulation 1205/2008]	Date of last revision	Conditional: at least one date of publication / date of creation / date of revision is required	dentificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date # Data type (and ISO19115 no.) 393. CI_Date	Described in ISO 19108 and ISO 8601
[Regulation 1205/2008]	Date of creation	Conditional: at least one date of publication / date of creation / date of revision is required	identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date # Data type (and ISO19115 no.) 393. CI_Date	Described in ISO 19108 and ISO 8601
[Regulation 1205/2008]	Lineage	Mandatory	dataQualityInfo/*/lineage/*/statement # Data type (and ISO19115 no.) CharacterString	Free text
[Regulation 1205/2008]	Spatial resolution	Mandatory if an equivalent scale or a resolution distance can be specified	identificationInfo[1]/*/spatialResolution/*/equivalentScale/*/denominator (equivalent scale) identificationInfo[1]/*/spatialResolution/*/distance (distance) # Data type (and ISO19115 no.) Integer (equivalent scale) Distance (distance)	-> positive integer (equivalent scale) -> number expressing the distance value and a unit of measure of the distance value (distance)
[Regulation 1205/2008]	Conformity/Specification	Mandatory for each conformity statement	dataQualityInfo/*/report/*/result/*/specification # Data type (and ISO19115 no.) 359. CI_Citation	The following properties are expected: -> Title (characterString and free text) -> Reference date (CI_Date): dateType: creation, publication or revision date: an effective date
[Regulation 1205/2008]	Conformity/Degree	Mandatory for each conformity statement	dataQualityInfo/*/report/*/result/*/pass # Data type (and ISO19115 no.) Boolean	-> true if conformant -> false if not conformant -> null (with nilReason = "unknown") if not evaluated

Implementation rule	Metadata Element Name	INSPIRE obligation/ condition/ note	ISO/TS 19139 path	Domain
[Regulation 1205/2008]	Conditions applying to access and use	Special values for unknown conditions or no applying conditions may be used	identificationInfo[1]/*/resourceConstraints/*/accessConstraints # Data type (and ISO19115 no.) MD_RestrictionCode identificationInfo[1]/*/resourceConstraints/*/otherConstraints # Data type (and ISO19115 no.) CharacterString identificationInfo[1]/*/resourceConstraints/*/useConstraints # Data type (and ISO19115 no.) MD_RestrictionCode identificationInfo[1]/*/resourceConstraints/*/otherConstraints # Data type (and ISO19115 no.) CharacterString	-> Codelist (strictly limited to the value defined in B.5.24 of ISO 19115) -> Free text or if the values "no conditions apply" or "conditions unknown" is used then an Anchor to the code list http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse in the Inspire Registry should be used. See also Annex D.2 in this document for the code list. -> Codelist (strictly limited to the value defined in B.5.24 of ISO 19115) -> Free text or if the values "no conditions apply" or "conditions unknown" is used then an Anchor to the codelist in the Inspire Registry should be used.
[Regulation 1205/2008]	Limitations on public access	Special value for no limitations may be used	identificationInfo[1]/*/resourceConstraints/*/accessConstraints # Data type (and ISO19115 no.) MD_RestrictionCode identificationInfo[1]/*/resourceConstraints/*/otherConstraints # Data type (and ISO19115 no.) Gmx:anchor	-> Codelist (strictly limited at the value defined in B.5.24 of ISO 19115) -> A code list value from the code list at http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess/ . See also Annex D.1 of this document for this code list.
[Regulation 1205/2008]	Responsible organisation/Responsible party	Mandatory for each responsible organisation	identificationInfo[1]/*/pointOfContact # Data type (and ISO19115 no.) 374. CI_ResponsibleParty	The following properties are expected: -> organisationName (characterString and free text) -> contactInfo (CI_Contact): address: electronicMailAddress [1..*] (characterString)
[Regulation 1205/2008]	Responsible organisation/Responsible party role	Mandatory for each responsible organisation	identificationInfo[1]/*/pointOfContact/*/role # Data type (and ISO19115 no.) CI_RoleCode	Codelist (see B.5.5 of ISO 19115)
[Regulation 1205/2008]	Metadata point of contact	Mandatory	Contact # Data type (and ISO19115 no.) 374. CI_ResponsibleParty	The following properties are expected: -> organisationName (characterString and free text) -> contactInfo (CI_Contact): address: electronicMailAddress [1..*] (characterString)
[Regulation 1205/2008]	Metadata date	Mandatory	dateStamp # Data type (and ISO19115 no.) Date	ISO 8601
[Regulation 1205/2008]	Metadata language	Mandatory	Language # Data type (and ISO19115 no.) LanguageCode (ISO/TS 19139)	Codelist (See ISO/TS 19139) based on alpha-3 codes of ISO 639-2. Use only three-letter codes from in ISO 639-2/B (bibliographic codes), The list of codes for the 24 official EU languages is: Bulgarian – bul Irish – gle Croatian – hrv Italian – ita Czech – cze Latvian – lav Danish – dan Lithuanian – lit Dutch – dut Maltese – mlt English – eng Polish – pol

Implementation rule	Metadata Element Name	INSPIRE obligation/ condition/ note	ISO/TS 19139 path	Domain
				Estonian – est Portuguese – por Finnish – fin Romanian – rum French – fre Slovak – slo German – ger Slovenian – slv Greek – gre Spanish – spa Hungarian – hun Swedish – swe The list of all the codes is defined at http://www.loc.gov/standards/iso639-2/ Regional languages also are included in this list.
[Regulation 1089/2010]	Coordinate Reference System	Mandatory to comply with [Regulation 1089/2010]	referenceSystemInfo # Data type (and ISO19115 no.) 186. MD_ReferenceSystem	To identify the reference system, referenceSystemIdentifier (RS_Identifier) shall be provided. RS_Identifier itself is a complex type (lines 206-207 and 208.1-208.2 from ISO 19115). At least the following element that is mandatory for ISO should be used (the multiplicity according to ISO 19115 is shown in parentheses): - 207. code [1] / domain value: free text TG Requirement 2 in INSPIRE Data specifications states that a URI identifier listed in a table provided there shall be used for referring to the Coordinate reference system. This table is provided as Annex D.5 of this document. If the code is given as an URI as shown above, the element codespace is not needed. The identifiers can be accessed via gmx:Anchor (see XML example). For regions outside of continental Europe, Member States may define suitable coordinate reference systems
[Regulation 1089/2010]	Temporal Reference System	Mandatory for compliance with [Regulation 1089/2010] only if a non-default temporal reference system (i.e. Gregorian Calendar or the Coordinated Universal Time) is used	referenceSystemInfo # Data type (and ISO19115 no.) 186. MD_ReferenceSystem	No specific type is defined in ISO 19115 for temporal reference systems. Thus, the generic MD_ReferenceSystem element and its referenceSystemIdentifier (RS_Identifier) property shall be provided. RS_Identifier itself is a complex type (lines 206-207 and 208.1-208.2 from ISO 19115). At least the following element that is mandatory for ISO should be used (the multiplicity according to ISO 19115 is shown in parentheses): - 207. code [1] / domain value: free text
[Regulation 1089/2010]	Encoding	Mandatory to comply with [Regulation 1089/2010]	distributionInfo/MD_Distribution/distributionFormat # Data type (and ISO19115 no.) 284. MD_Format	This is a complex type (lines 285-290 from ISO 19115). At least the following elements that are mandatory for ISO should be used (the multiplicity according to ISO 19115 is shown in parentheses): - 285. name [1] / domain value: free text - 286. version [1] / domain value: free text Content for name could also be taken from INSIPRE Registry using the code list available here: http://inspire.ec.europa.eu/media-types/ and can be accessed via gmx:Anchor (see XML example).
[Regulation 1089/2010]	Character Encoding	Conditional for dataset and dataset series; Mandatory if NOT using	identificationInfo[1]/*/characterSet # Data type (and ISO19115 no.) MD_CharacterSetCode	CodeList (see B.5.10 of ISO 19115)

Implementation rule	Metadata Element Name	INSPIRE obligation/ condition/ note	ISO/TS 19139 path	Domain
		standard UTF-8 encoding		
[Regulation 1089/2010] amended by [Regulation 1253/2013]	Spatial representation type	Mandatory to comply with [Regulation 1089/2010]	identificationInfo[1]/*/spatialRepresentationType # Data type (and ISO19115 no.) MD_SpatialRepresentation TypeCode	CodeList (see B.5.26 of ISO 19115), following INSPIRE Data specifications only **vector**, **grid** and **tin** should be used.
[Regulation 1089/2010]	Topological consistency	Mandatory for compliance with [Regulation 1089/2010]; Conditional for dataset and dataset series: mandatory if the data set includes types from the Generic Network Model and does not assure centreline topology (connectivity of centrelines) for the network	dataQualityInfo/DQ_DataQuality/report/ # Data type (and ISO19115 no.) 115. DQ_TopologicalConsistency dataQualityInfo/DQ_DataQuality/report # Data type (and ISO19115 no.) 115. DQ_TopologicalConsistency	->**Topological Consistency-Quantitative results** DQ_TopologicalConsistency is a forming of the abstract complex type DQ_Element. See B.2.4.3 in ISO 19115:2003 for further information. The following ISO 19115 elements are the corresponding ones to express quantitative results of the data quality evaluation as given in INSPIRE Data specifications sections 8.3.2 which in fact focus on ISO 19157: - 100. nameOfMeasure [0..*]: name of the test applied to the data / domain value: free text - 103. evaluationMethodType [0..1]: type of method used to evaluate quality of the dataset/ domain value: DQ_EvaluationMethod TypeCode - 104. evaluationMethodDescription [0..1]: description of the evaluation method / domain value: free text - 106. dateTime [0..*]: date or range of dates on which a data quality measure was applied / domain value: DateTime (ISO 19103) - 107. result [1..2]: value (or set of values) obtained from applying a data quality measure or the outcome of evaluating the obtained value (or set of values) against a specified acceptable conformance quality level / domain value: DQ_Result (abstract) - 133. DQ_QuantitativeResult, consisting of - 137. value [1..*]: quantitative value or values, content determined by the evaluation procedure used / domain value: Record (ISO 19103) Due to making use of DQ_QuantitativeResult subset there is a mandatory element in ISO 19115 to be considerer too: - 135. valueUnit [1] ->**Topological Consistency-Descriptive results** DQ_TopologicalConsistency is a forming of the abstract complex type DQ_Element. See B.2.4.3 in ISO 19115:2003 for further information. To provide the descriptive results of Topological consistency evaluation DQ_ConformanceResult containing the following elements should be used (the multiplicity according to ISO 19115 is shown in parentheses): - 130. specification [1..1]: citation of product specification or user requirement against which data is being evaluated / domain value: CI_Citation - 131. explanation [1..1]: explanation of the meaning of conformance for this result / domain value: free text - 132. pass [1..1]: indication of the conformance result / domain value: Boolean

Table 3-5 (cont.)

All the mandatory or conditional metadata elements (Tab. 3-5) are recommended for all datasets or datasets series included in all INSPIRE annex(s) theme(s) (or Thematic Category) (Tab. 3-6).

Tab. 3-6. INSPIRE Annex(s), INSPIRE Spatial Data Themes (theme code), and Topic category.

INSPIRE	**Spatial Data Themes** (Thematic Category)	**Code**	**Topic Category**
Annex I	01. Coordinate reference systems	RS	-
	02. Geographical grid systems	GG	-
	03. Geographical names	GN	Location
	04. Administrative units	AU	Boundaries
	05. Addresses	AD	Location
	06. Cadastral parcels	CP	planningCadastre
	07.Transport networks	TN	Transportation
	08. Hydrography	HY	inlandWaters
	09. Protected sites	PS	Environment
Annex II	01. Elevation	EL	Elevation
	02. Land cover	LC	imageryBaseMapsEarthCover
	03. Orthoimagery	OI	imageryBaseMapsEarthCover
	04. Geology	GE	geoscientificInformation
Annex III	01. Statistical units	SU	Boundaries
	02. Buildings	BU	Structure
	03. Soil	SO	geoscientificInformation
	04. Land use	LU	planningCadastre
	05. Human health and safety	HH	Health
	06. Utility and governmental services	US	utilitiesCommunication
	07. Environmental monitoring facilities	EF	Structure
	08. Production and industrial facilities	PF	Structure
	09. Agricultural and aquaculture facilities	AF	Farming
	10. Population distribution – demography	PD	Society
	11. Area management/ restriction/ regulation zones and reporting units	AM	planningCadastre
	12. Natural risk zones	NZ	geoscientificInformation
	13. Atmospheric conditions	AC	climatologyMeteorologyAtmosphere
	14. Meteorological geographical features	MF	climatologyMeteorologyAtmosphere
	15. Oceanographic geographical features	OF	Oceans
	16. Sea regions	SR	Oceans
	17. Bio-geographical regions	BR	Biota
	18. Habitats and biotopes	HB	Biota
	19. Species distribution	SD	Biota
	20. Energy resources	ER	Economy
	21. Mineral resources	MR	economy

This technical specification (ISO/TS19139:2007), also recommends, INSPIRE specific-metadata (optional) element for each INSPIRE annex theme(s) or Thematic Category (Tab. 3-7). The required INSPIRE metadata elements for spatial datasets and datasets series, for metadata conformity with INSPIRE, is provided in the Tab. 3-8, as well as ISO/TS 19139 path (namespace) and domain value(s).

Tab. 3-7. Mapping table of theme-specific INSPIRE metadata with INSPIRE annex themes (adapted from ISO/TS19139).

INSPIRE theme specific metadata element	Definition	**Which annex theme is involved?** (see Tab. 3-6 for legend)																																	
		RS	GG	GN	AU	AD	CP	TN	HY	PS	EL	LC	OI	GE	SU	BU	SO	LU	HH	US	EF	PF	AF	PD	AM	NZ	AC	MF	OF	SR	BR	HB	SD	ER	MR
Maintenance information	Information about the scope and frequency of updating	-	-	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	x	x	x	x
Spatial representation information	Digital representation of spatial information in the dataset	-	-								x													-											
Supplemental information	Any other descriptive information about the dataset	-	-								x													-											
Process step	Information about an event or transformation in the life of a dataset including the process used to maintain the dataset	-	-								x		x											-											
Data source	Information about the source data used in creating the data specified by the scope	-	-								x		x											-											
Browse graphic information	Graphic that provides an illustration of the dataset (should include a legend for the graphic)	-	-								x		x											-											
Image description	Information about an image's suitability for use	-	-										x											-											
Content information	description of the content of a dataset	-	-													x								-											
Digital transfer options information	Technical means and media by which a resource is obtained from the distributor	-	-						x		x		x											-											
Identification - Extent	In addition to the Geographic bounding box the following element should be used to provide a common "name" for the extent	-	-						x															-											
Data Quality – QUANTITATIVE RESULTS	Several, theme specific aspects of Data Quality - See corresponding Data Specification for further information	see separate table for details on data quality mapping (Tab. 3-9)																																	

Tab. 3-8. Overview of the required theme-specific metadata elements (optional) from INSPIRE Data Specifications (ISO/TS19139 adapted)

Implementation rule	**Metadata Element Name**	**INSPIRE obligation/ condition/ note**	**Recommended for themes**	**ISO/TS19139 path**	**Domain**
INSPIRE Data specifications, sections 8.3.1	Maintenance information	Optional	**all**	identificationInfo[1]/MD_DataIdentification/resourceMaintenance/ # Data type (and ISO19115 no.) 142. MD_MaintenanceInformation	This is a complex type (lines 143-148 from ISO 19115). At least the following element should be used (the multiplicity according to ISO 19115 is shown in parentheses): - 143. maintenanceAndUpdateFrequency [1]: frequency with which changes and additions are made to the resource after the initial resource is completed / domain value: MD_MaintenanceFrequencyCode In addition the following elements are recommended, but in contrast to ISO each of them should not appear multiple but single only: - 146. updateScope [0..*]: scope of data to which maintenance is applied / domain value: MD_ScopeCode - 148. maintenanceNote [0..*]: information regarding specific requirements for maintaining the resource / domain value: free text
INSPIRE Data specifications, sections 8.3.x	Spatial representation information	Optional	**Elevation**	spatialRepresentationInfo/ # Data type (and ISO19115 no.) 156. MD_SpatialRepresentation	MD_SpatialRepresentation is an abstract complex type and has to be expressed as MD_GridSpatialRepresentation, MD_Georectified, MD_Georeferenceable or MD_VectorSpatialRepresentation. See B.2.6 in ISO 19115:2003 for further information.
INSPIRE Data specifications, sections 8.3.x	Supplemental information	Optional	**Elevation**	identificationInfo[1]/MD_DataIdentification/supplementalInformation # Data type (and ISO19115 no.) CharacterString	Free text
INSPIRE Data specifications, sections 8.3.x	Process step	Optional	**Elevation; Orthoimagery**	dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/ # Data type (and ISO19115 no.) 86. LI_ProcessStep	This is a complex type (lines 87-91 from ISO 19115). The description (87., free text) property shall be provided. Comments: ISO 19115 lists several elements which build LI_ProcessStep. For the purpose of theme-specific metadata according to the INSPIRE Data specifications the element listed above is sufficient. Note that the path for dataQualityInfo/DQ_DataQuality/lineage/ will already exist in metadata because of being used for carrying information about lineage itself (see 2.7.1). Therefore an addition of these information into the same entity of LI_Lineage may be useful.
INSPIRE Data specifications, sections 8.3.x	Data source	Optional	**Elevation; Orthoimagery**	dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/ # Data type (and ISO19115 no.) 85. LI_Source	This is a complex type (lines 93-98 from ISO 19115). Either the description (93., free text) or the sourceExtent (97., EX_Extent) elements shall be provided.
INSPIRE Data specifications, sections	Browse graphic information	Optional	**Elevation; Orthoimagery**	identificationInfo[1]/MD_DataIdentification/graphicOverview	This is a complex type (lines 49-51 from ISO 19115). The following element is mandatory (the

Implementation rule	Metadata Element Name	INSPIRE obligation/ condition/ note	Recommended for themes	ISO/TS19139 path	Domain
8.3.x				# Data type (and ISO19115 no.) 48. MD_BrowseGraphic	multiplicity according to ISO 19115 is shown in brackets): - 49. filename [1]: name of the file that contains a graphic that provides an illustration of the dataset / domain value: free text
INSPIRE Data specifications, sections 8.3.x	Image description	Optional	**Orthoimagery**	contentInfo/ # Data type (and ISO19115 no.) 243. MD_ImageDescription	This is a complex type (lines 244-255 and 249-242 from ISO 19115). At least the following element should be used: - 248. cloudCoverPercentage [1]: area of the dataset obscured by clouds, expressed as a percentage of the spatial extent/ domain value: Real ISO 19115 itself demands two mandatory elements in MD_ImageDescription: - 240. attributeDescription [1] - 241. contentType [1]
INSPIRE Data specifications, sections 8.3.x	Content information	Optional	**Buildings**	contentInfo/ # Data type (and ISO19115 no.) 233. MD_FeatureCatalogueDescription	This is a complex type (lines 234-238 from ISO 19115). Data specification on Buildings does not give a minimum element. ISO 19115 itself demands two mandatory elements in MD_ FeatureCatalogueDescription: - 236. includedWithDataset [1] - 238. featureCatalogueCitation [1]
INSPIRE Data specifications, sections 8.3.x	Digital transfer options information	Optional	**Hydrography; Elevation; Orthoimagery**	distributionInfo/MD_Distribution/transferOptions/ # Data type (and ISO19115 no.) 274. MD_DigitalTransferOptions	This is a complex type (lines 275-278 from ISO 19115). At least the following elements should be used (the multiplicity according to ISO 19115 is shown in parentheses): For Elevation and Orthoimagery: - 275. unitsOfDistribution [0..1]: tiles, layers, geographic areas, etc., in which data is available / domain value: free text - 278. offLine [0..1]: information about offline media on which the resource can be obtained / domain value: MD_Medium For Hydrography: - 276. transferSize [0..1]: estimated size of a unit in the specified transfer format, expressed in megabytes. The transfer size is > 0.0/ domain value: Real
INSPIRE Data specifications, sections 8.3.x	Extent	Optional	**Hydrography**	identificationInfo[1]/MD_DataIdentification/extent # Data type (and ISO19115 no.) 334. EX_Extent	This is a complex type (lines 335-338 from ISO 19115). In addition to the Geographic bounding box (see 2.3.8) the following element should be used to provide a common "name" for the extent (the multiplicity according to ISO 19115 is shown in parentheses): - 335. description [0..1]: spatial and temporal extent for the referring object/ domain value: free text

Table 3-8 (cont.)

Metadata elements concerning data quality (DQ), based on ISO19157:2013 (concepts, namespace or package, and domain values) are based on INSPIRE implementation rules (Tab. 3-10), as well as the requirements for each specific INSPIRE spatial data themes (Tab. 3-9).

Tab. 3-9. Overview table of theme specific INSPIRE metadata elements concerning data quality (DQ) (ISO/TS19139 adapted).

INSPIRE theme specific metadata element for Data quality		**Which annex theme is involved?** (see Tab. 3-6 for legend)																																	
		RS	GG	GN	AU	AD	CP	TN	HY	PS	EL	LC	OI	GE	SU	BU	SO	LU	HH	US	EF	PF	AF	PD	AM	NZ	AC	MF	OF	SR	BR	HB	SD	ER	MR
Completeness	Commission	-	-		x	x		x	x	x	x	x			x	x		x		x				-											
	Omission	-	-	x	x	x	x	x	x	x	x	x	x		x	x	x	x		x		x	x	-					x						
Logical Consistency	Conceptual consistency	-	-	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	x	x	x	x
	Domain consistency	-	-	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	x	x	x	x
	Format consistency	-	-					x			x	x								x				-											
Positional Accuracy	Absolute or external accuracy	-	-	x	x	x	x	x	x	x	x	x			x	x		x	x					-					x	x					
	Relative or internal accuracy	-	-						x			x												-											
	Gridded data positional accuracy	-	-								x		x											-											
Temporal Quality	Temporal consistency	-	-									x												-											
	Temporal validity	-	-									x			x				x			x	x	-											
Thematic Accuracy	Thematic classification correctness	-	-					x				x			x			x				x	x	-											
	Non-quantitative attribute correctness	-	-			x		x	x			x						x						-											
	Quantitative attribute accuracy	-	-						x			x							x					-											
Usability	---															x													x	x					

Tab. 3-10. Overview (example: DQ_Commission) of the required the data quality elements (ISO/TS19139 adapted).

Implementation rule	Metadata Element Name	INSPIRE obligation / condition / note	Recommended for themes	ISO/TS19139 path	Domain
ISO19157	**DQ_Element** (for example): DQ_Commission	Optional	**Administrative units; Addresses; Transport networks; Hydrography; Protected sites; Elevation; Land cover; Statistical units; Buildings; Land use; Utility and governmental services** [See Tab. 3-9]	dataQualityInfo/DQ_DataQuality/report/ # Data type (and ISO19115 no.) 90. DQ_Element	**-> Data Quality-Quantitative results** DQ_Element is an abstract complex type and has to be expressed by a corresponding DQ_Commission. The following ISO 19115 elements are the corresponding ones to express quantitative results of the data quality evaluation as given in INSPIRE Data specifications sections 8.3.2 which in fact focus on **ISO19157**: - 100. nameOfMeasure [0..*]: name of the test applied to the data / domain value: free text - 103. evaluationMethodType [0..1]: type of method used to evaluate quality of the dataset/ domain value: DQ_EvaluationMethod TypeCode - 104. evaluationMethodDescription [0..1]: description of the evaluation method / domain value: free text - 106. dateTime [0..*]: date or range of dates on which a data quality measure was applied / domain value: DateTime (ISO 19103) - 107. result [1..2]: value (or set of values) obtained from applying a data quality measure or the outcome of evaluating the obtained value (or set of values) against a specified acceptable conformance quality level / domain value: DQ_Result (abstract) - 133. DQ_QuantitativeResult, consisting of - 137. value [1..*]: quantitative value or values, content determined by the evaluation procedure used / domain value: Record (ISO 19103) Due to making use of DQ_QuantitativeResult subset there is a mandatory element in ISO 19115 to be considerer too: - 135. valueUnit [1] **->Data Quality-Descriptive results** DQ_Element is an abstract complex type and has to be expressed by a corresponding DQ_Commission. See B.2.4.3 in ISO 19115:2003 for further information. To provide the descriptive results of Topological consistency evaluation DQ_ConformanceResult containing the following elements should be used (the multiplicity according to ISO 19115 is shown in parentheses): - 130. specification [1..1]: citation of product specification or user requirement against which data is being evaluated / domain value: CI_Citation - 131. explanation [1..1]: explanation of the meaning of conformance for this result / domain value: free text - 132. pass [1..1]: indication of the conformance result / domain value: Boolean

SDQE requirements imply the introduction of mandatory and optional attributes proposed as a support to generation of (in)direct quality indicators in external quality evaluation (Tab. 3-11).

Tab. 3-11. Data characteristics and suggestion of related quality indicators (Task 5.5 ECOPOTENTIAL project).

Main Characteristics	Quality elements/ Indicators	Description/Value domain/comparison type
Definition	Typology	Topic category defining the main dataset theme/ Text string / String comparison
Coverage	Spatial extent	Bounding box defining spatial coverage/ Geometric bounding box/ **% of cover**
	Temporal extent	Time interval defining temporal coverage/ Start date – End date/ **% of cover**
Lineage	Lineage description in metadata	Description of data production methods and overall quality/ Text-character-string/ **Boolean comparison**
Precision	Spatial scale	Equivalent scale or spatial resolution defining the level of detail / Integer – Double/ **Intersection test**
Legitimacy	Producer recognition	Data producer recognition type/ Text-string/ **String comparison**
Accessibility	Access and use constraints	Conditions applying to access and use (Directive 2007/2/EC) / Text-string/ **String comparison**
	File format	Distribution file format/ Text-string/ **String comparison**
	Reference systems	Reference systems/ Text-string/ **String comparison**

3.3.2 Spatial metadata fulfilment, catalogue and (meta)data management

Spatial metadata facilitate the communication, selection, discovery and share relevant data on thematic, open, distributed, dynamic and multidisciplinary communities associated to project and local/global initiatives. All pre-existing and processing (meta)data should considers global and thematic (meta)data specification standards. Therefore, the potentialities of the metadata require processes of fulfilment, organization and management, as well as transforming the pre-existing catalogues. The introduction of quality fields on selected metadata profile aims to perform spatial data fitness for use quality evaluation.

Spatial metadata facilitate the communication, selection, discovery and share relevant data on thematic, open, distributed, dynamic, multifaceted and multidisciplinary communities associated to project and local/global initiatives. All pre-existing and processing (meta)data should consider global and thematic (meta)data specification standards. Therefore, the potentialities of the metadata require processes of fulfilment, organization and management, as well as transforming the pre-existing catalogues. The introduction of quality fields on selected metadata profile aims to perform spatial data fitness for use quality evaluation.

The proposal of routines and metadata fitness for use implies the development of (i) user´s requirements analysis and product specification (focused on user's requirements, spatial data quality elements specification/utility…) and (ii) quality evaluation routines development (focused on implementation of storylines "Theoretical framework – WP2" case-study Peneda-Gerês and others…). According to the previous section, it is critical to improve spatial metadata fulfilment along the metadata life cycle. This major and central challenge implies the development of: (i) individual and collective capacity building related to metadata relevance, metadata profile standards and metadata core contents aiming to perform interoperability, share and discovery of critical (meta)data; (ii) processes of (semi)automatic fulfilment mandatory metadata with technical guidance process documentation, and implementation of (meta)data quality control and assurance at information system/infrastructure management and organizational quality governance framework; (iii) tools that improve and support metadata fulfilment and promote metadata transformation, harmonization and metadata interoperability related to metadata enrichment/relevance associated to thematic knowledge networks management.

Given the relevance and criticalness of metadata, it is fundamental to have catalogues with services accessible through the Internet that can be used by the geospatial community to facilitate data communication and sharing. Key factors for these services are interoperability and open standards in order to promote information harmonization and sharing as well as to enable the interoperability and integration of information systems. In the case of geospatial data, the Open Geospatial Consortium (OGC) standards have become the core standards (Lopez-Pellicer et al., 2011), namely the Catalogue Services for the Web (CSW) (OGC, 2007), which offers an interface to publish metadata on geographic information, and to provide mechanisms to search, access and maintain metadata information based on communication protocols/standards. The provision of services to support discovery and access to geographic information resources is a critical issue as they can be used to allow users to verify the availability of appropriate data fitting their requirements to instruct a specific analytical or modelling process, and ultimately retrieve the respective datasets. In this context, it is fundamental to develop tools to support user oriented discovery of data based on metadata to assess the existence of datasets that comply user's needs, but also, to implement catalogue services that supports different standard metadata profiles and the conversion between profiles in order to encompass a variety internal and external data sources.

Task 5.7 aims: (i) the creation of a database to guarantee robust and reliable exchange of interoperable data [D5.7]; (ii) Provide a framework of interoperable data to the modelling WPs based on the model requirements and setup of data flows from relevant data sources (defined by Task 5.1). Provision of a well-structured database framework (following the data model/format specifications defined in Task 5.6) to robustly and reliably store datasets, to provide data as a service to the project. The database will be optimized for large dataset storage as well as maximizing access performance. Web portal facilities integrating the database will be available for a successful access and analysis (in cooperation with WP10; interfaces to connect to the GEO DAB via the ECOPOTENTIAL Virtual Laboratory Platform will be supplied).

Although Task 5.5 falls within *WP5: In situ Monitoring Data*, the ECOPOTENTIAL project has organized its workflow around storylines. Storylines are narratives supporting customized workflows aimed to provide estimates of indicators of specific benefits provided to one or more protected areas. The Storylines include specific modelling frameworks or "pipelines" which are fed by derived variables (Essential Variables - EVs) and ancillary variables, both obtained from EO and/or in-situ data. All (or most) storylines have provided schematic views of their workflows in the form of mind maps (WP7) and as workflows/schemes for deliverable D2.2. The implementation of a given storyline in a given protected area is ensured by one or more scientific partners and by protected area managers. Therefore, for Task 5.5 implementation, the Storylines provide a suitable context for testing methodologies and tools for data (and product) quality assessment and management in ECOPOTENTIAL Project.

The discussions on Remote sensing (RS) variable selection and on the links between each storyline (e.g. Schematic work flow of the Peneda-Gerês storyline [M7]) and the strategy for Task 5.5 have highlighted the importance of detailing the data processing-modelling pipelines for each of the indicators proposed in the storyline. In the case of the habitat quality indicator, we have two pipelines and so in total we will have five pipelines. For each pipeline, this would mean identifying as precisely as possible: (i) which indicator will be estimated?; (ii) which EO and in situ data will be used?; (iii) which EVs will be processed and fed to the model?; (iv) which model will be used?; (v) the main features of model outputs (e.g. spatial and temporal resolutions)?.

These exercises will be very useful for our work in task 5.5 (evaluating data quality along pipelines), to setup all our pipelines in the new cluster, to plan upcoming field surveys and communication with partners with whom we share model development or data processing. We suggest to develop this exercise based on the most recent versions of the mind maps (WP7) and of the work flows outlined schematically in deliverable D2.2 (Fig. 3-22).

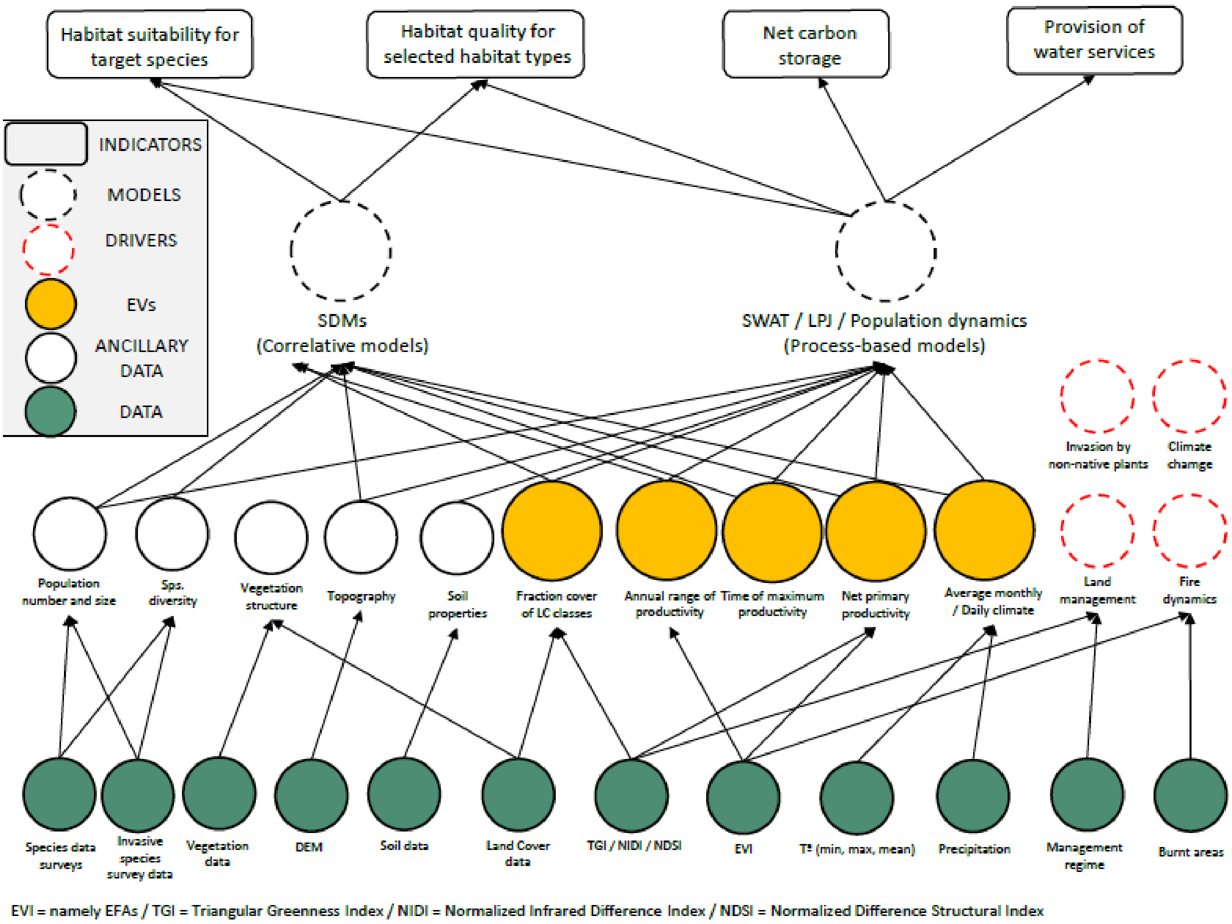


Fig. 3-22. Schematic work flow of the Peneda-Gerês storyline [M7] (Source: deliverable D2.2).

Users should identify the management broader or specific application contexts for the outputs of their special analysis models (be they process-based model and correlative model ...). For each application context model (related to WP2, conceptual framework for Storyline(s) implementation), the user identifies the data (EO/RS and in situ datasets or datasets series, ancillary data, and essential variables -EVs) to inform a particular model, and define quality aspects that would like to see reflected in your results (expected quality, depending on user requirements, model, or legal requirements). The user should describe the application context (rationale), select the quality indicators that he considers most relevant in terms of his application context versus analysis model and/or require normative or user-defined quality values for these quality indicators selected.

The possibility to consult (related to task 5.7) metadata catalogues (related to task 5.3) to identify, know and select data or datasets to inform analysis models (WP6) aims to communicate data (PA) and quality requests supported by the user´s needs of the spatially explicit models or technical-legal aspects allowing to verify availability and conditions of access to them, as well as to identify data gaps (related to task 5.1). In this context it is important to establish the user data and data quality requirements for a given application context and the legal / regulatory requirements to comply with a particular technical specification / guidance (WP6 / WP7). In this sense, there is a great interest for users towards the completeness of the metadata catalogues, the possibility of being provided with consultable and interoperable metadata catalogues, allowing their consultation and readings, translating into information that assists them in their work routines, facilitating research and reducing the time spent with these recurring tasks.

The ability to share/communicate data needs to inform a particular model of special analysis and generate any indicator (e.g. ecosystem service) for a given protected area (PA) can help identify possible joint work between different teams. These purposes involve paying particular attention to the challenge of documenting metadata in a harmonized way (implementing standards and associated implementation rules) on the production and maintenance/updating of the data (over the data life cycle) (e.g. time date series). This assumption makes evident the need to work on the definition and implementation of technical guidelines for creation of metadata, management guidelines and (meta)data sharing, considering the (semi)automation of quality management processes and procedures, facilitating the operationalization and improvement of the results of individual and collective analysis, communication and decision in knowledge networks.

4. THEmatic Metadata-based and fItness for use Spatial data quality Evaluation

The main objective of the ***THEmatic Metadata-based and fItness for use Spatial data quality Evaluation*** (ThemisE[35] platform) is to develop and implement user-oriented spatial data quality evaluation routines and procedures (data quality management) for the assessment of the internal and external quality of pre-existing data, based on quality measures extracted from (spatial) metadata, accessible as a web-based application based on open-sources technologies. Within ECOPOTENTIAL's scope, these tools are important to support the quality-driven identification/selection of relevant data or the identification of data quality gaps enabling the planning of targeted data collection.

The provision of tools to support the searching and selection of relevant data is a critical issue given that there is an increasing amount of spatial data production, handling and sharing, with different sources, different frequency of acquisition, different spatio-temporal scales and levels of accuracy, different processing methods or techniques. This fact leads to many challenges considering that spatial data is used in very different application contexts with data often used with purposes other than producer' intended ones. There is a risk of misusing and/or misinterpretation of data by users that can cause misleading results. Therefore, data quality can be valuable instrument to consider in the identification process of datasets that satisfy the requirements of a particular application for a specific user.

In this context, the selected approach for the platform development is based on the methodological framework developed by Honrado et al., 2011a and Pôças et al., 2014. This approach allows evaluating the fitness for use of spatial datasets centred on user requirements. This is accomplished by analysing the similarities between the user's requirements and the data sets characteristics based on a set of quality indicators as detailed by metadata. This approach intends to support the user to evaluate which datasets fulfil his/her expectations and to identify the more appropriate datasets to be used to solve his/her problem.

For the purpose of this task, the implemented platform, besides being based on the previously mentioned methodology, it includes some new advancements, in particular, the possibility to execute simultaneous quality-driven searching/identification of several datasets necessary to a certain application context through the definition of different sets of expected values for quality indicators. At the same time, it allows to evaluate metadata quality of datasets regarding required elements and specific standard profile(s).

Given these considerations, the following sub-sections will present: (i) the framework overview for user-oriented quality evaluation; (ii) the platform functionalities and user's requirements; (iii) the logical and technological architecture; and (iv) the implementation and functionalities of the ThemisE platform.

4.1 ThemisE platform framework overview

The development of ThemisE platform to support the evaluation and identification of relevant data for specific user´s application contexts is based on the specification of user-oriented quality evaluation routines

[35] **Themis** /ˈθiːmɪs/ (Greek: Θέμις) is an ancient Greek Titaness. She is described as "[the Lady] of good counsel", and is the personification of divine order, fairness, law, natural law, and custom. Her symbols are the Scales of Justice, tools used to remain balanced and pragmatic. *Themis* means "divine law" rather than human ordinance, literally "that which is put in place", from the Greek verb *títhēmi* (τίθημι), meaning "to put". To the ancient Greeks she was originally the organizer of the "communal affairs of humans, particularly assemblies" (https://en.wikipedia.org/wiki/Themis).

and procedures for the assessment of the quality of pre-existing data based on (spatial) metadata. In the case of the present project, two types of evaluation were foreseen: (i) one of the evaluations is centred on the comparison of the characteristics of the dataset, as detailed in metadata by the producer, with the required elements according to a predefined standard profile; (ii) the other evaluation is an external quality evaluation that, as mentioned previously, is based on determining the level of similarity between the characteristics of the data as detailed by metadata and a set of user-defined values of quality indicators that describe the user's requirements for a given application context.

From a conceptual perspective, the proposed methodological framework can be described as a set of steps performed by the user or by the platform for a certain application context (Fig. 4-1). One of the steps, to be executed by the user, consists in selecting the metadata catalogues as sources of the datasets that will be evaluated. The other step to be performed by the user consists in specifying the thematic category and the expected data quality values of available quality indicators for each dataset that is targeted for search, and defining also the quality indicators that are critical factors[36] or to be used as filters. This specification can be done for different targeted datasets that are relevant for the given application context. This approach aims to provide a high degree of freedom in adjusting quality criteria in relation to the application context, according to various user's needs for multiple datasets and, based on its expertise. The other steps are performed internally by the platform.

One of the steps involves the filling of a matrix with the quality indicators values for each dataset by inspection of metadata retrieved from catalogues. Each of these quality matrices is then used in the evaluation procedure to generate a final external quality matrix by comparing expected and datasets quality values. This is processed indicator by indicator (pairwise comparisons) for each user's defined targeted datasets, to determine which quality indicators are conformant or non-conformant. Note that during this process, some datasets retrieved from metadata catalogues are excluded from the evaluation process due to not comply with the user's defined filters. Simultaneously, the evaluation process generates another matrix for each dataset regarding the required metadata elements related to thematic categories of a specific metadata standard profile(s), where each matrix element has a value corresponding to the conformity or non-conformity for each metadata element/thematic category. Then, two final fitness values are calculated for each catalogue(s) datasets based on the previous matrices to allow the identification of the datasets that better fulfil the user´s requirements for the application context under consideration.

[36] Critical factors allow the user to identify important quality indicators that if not verified (non-conformant) will make the dataset invalid (unfit) for the specified application context.

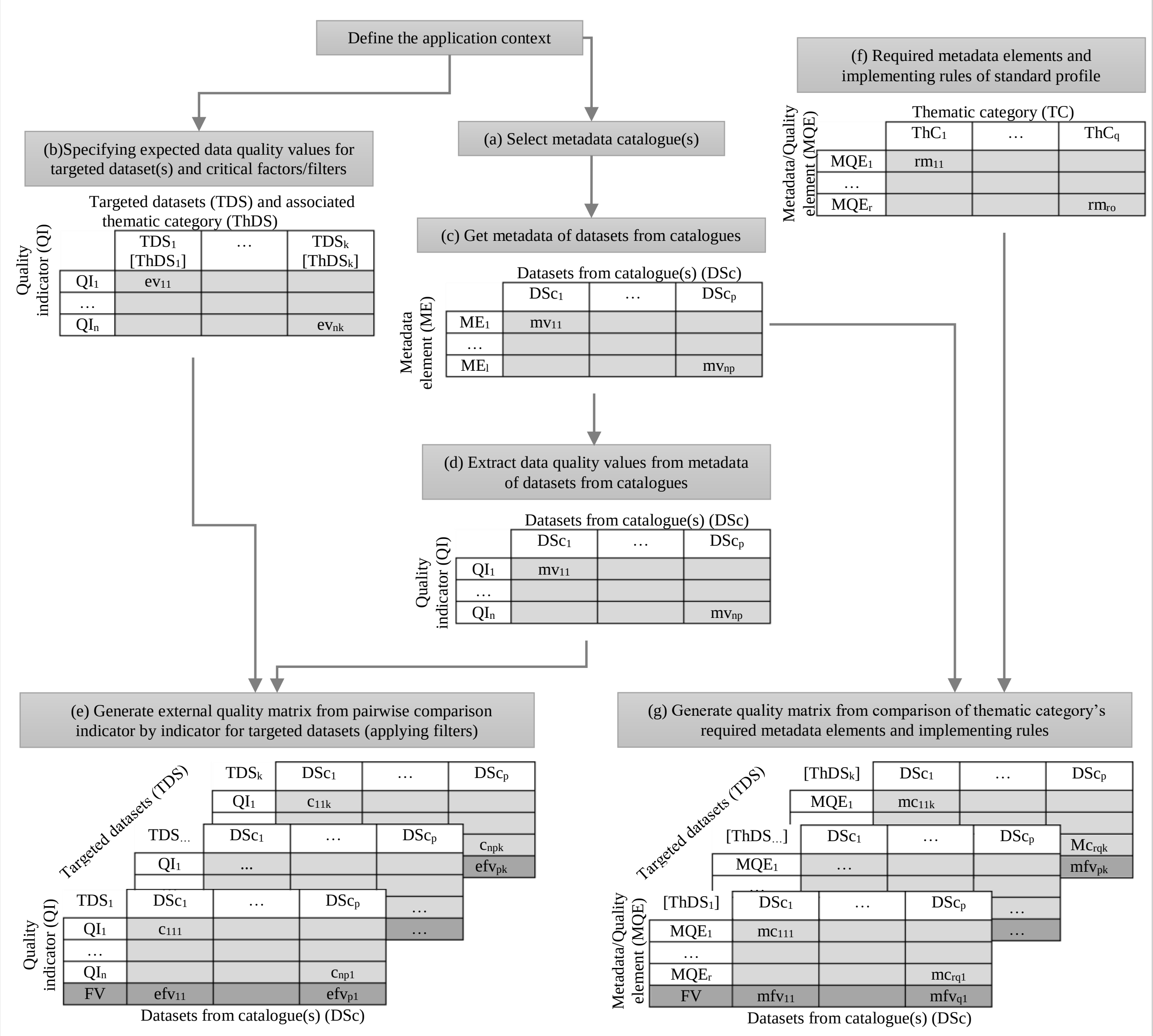


Fig. 4-1. General workflow of the proposed user-oriented quality evaluation framework.

The setting of the expected quality indicators values has to be done by the user considering the specificity of each targeted dataset and the application context (step b in Fig. 4-1). In this framework, the characteristics of quality indicators were based on the proposal of Pôças et al., 2014, where each quality indicator is based on a single metadata field corresponding to one quality characteristics, allowing that data quality values can be filled in, directly from characteristics detailed in the dataset's metadata retrieved from database catalogues (step d in Fig. 4-1). With respect to the available options to use when multiple values are defined for a quality indicator, a more flexible approach has been implemented to allow combining multi-level values using two connectors corresponding to logical connectives of conjunction and disjunction. For some quality indicators, an additional parameter, identified as cover, can be set to define the minimum expected percentage of coverage necessary to have a conformant result.

The generation of the external quality matrix (step e in Fig. 4-1) comprises the comparison between expected and dataset quality values for each metadata record retrieved from the catalogue(s). The comparison is made indicator by indicator for each targeted dataset (corresponding to a matrix element) and is based on predefined rules on how the pairwise comparison is performed for each quality indicator. The comparison result is a logical value corresponding to the conformity or non-conformity of the dataset metadata regarding the user's expected value(s). In order to offer a framework that is applicable to a wide variety of application contexts, the comparison between expected and dataset quality values is grounded on the rule-based system proposed in Honrado et al., 2011a with a new option to filter datasets. The system is configurable through the specification of a set of parameters for each quality indicator: the data type of the variable and the associated comparison method that will be used to compare the expected values with metadata values of datasets; the element of the metadata profile from which to extract values used in comparisons; options that allow to define if a quality indicator can be used as a critical factor or as a filter; and the possibility to specify if comparisons are made based on a percentage of cover (for spatial or temporal overlap). As an example, Tab. 4-1 presents some of the quality indicators used in the tests carried out during the development of the *ThemisE* platform. Nevertheless, we emphasize that *ThemisE* is extendable in order to include other quality indicators that might be relevant to use in other application contexts.

Tab. 4-1. Description of quality indicators for external quality evaluation.

Quality indicator	Data type [Comparison method]	Element from metadata profile	Options
Topic category	Predefined list of values [String comparison]	Topic category	Critical factor (optional)
Spatial scale	Scale defined as interval of numbers [Intersection test]	Spatial resolution	Critical factor (optional)
Spatial extent	Geographic bounding-box [overlapping percentage]	Geographic bounding box	Critical factor (optional) Filter (optional) Percentage of cover
Temporal extent	Time interval [overlapping percentage]	Temporal extent	Critical factor (optional) Filter (optional) Percentage of cover
Lineage	Free text or regular expression [String comparison]	Lineage description	Critical factor (optional)
Representation type	Predefined list of values [String comparison]	Distribution format (Distribution)	Critical factor (optional) Filter (optional)

The last step of the external quality evaluation process consists in the calculation of a ***fitness for use*** value for each pair of datasets from catalogue(s) and targeted datasets corresponding to the percentage of conformant quality indicators (row FV of matrices in step e of Fig. 4-1). Based on this fitness value and the indicators defined as critical factors, the evaluated datasets are classified as: (i) **unfit**, if at least one critical indicator is non-conformant, (ii) **partially fit,** if there is no critical indicator but at least one non-conformant indicator, and, (iii) **fit,** in the case of all indicators are conformant.

The quality evaluation regarding the required elements and implementing rules for metadata for a predefined standard profile comprises two steps. The first step encompasses the specification of a matrix that includes the common metadata elements of a specific profile according to defined implementing rules of all thematic categories and the required quality elements for each thematic category (step f in Fig. 4-1). During the implementation and testing of the *ThemisE* platform, the quality matrix was defined with mandatory or conditional metadata elements for metadata compliance (with implementation rule Regulation 1205/2008, for the INSPIRE standard, based on the implementation rules of ISO/TS19139:2007), as well as the optional and recommended quality elements (*DQ_Elements*) for each INSPIRE thematic category (data quality elements to be evaluated according to the corresponding thematic category) (Tab. 3-9). The second step, performed by the platform, includes the generation of a metadata/quality elements matrix for each targeted dataset according to its thematic category, which comprises the verification of compliance of the recommended metadata/quality elements defined in the previous step for each metadata record retrieved from catalogue(s). Lastly, a global compliance value is calculated for each pair of datasets from catalogue(s) and targeted datasets corresponding to the percentage of fulfilment of metadata/quality elements defined in its thematic category (Tab. 4-2).

Regarding the entire evaluation process, it should be noted that thematic categories are a central element throughout the process. These can be considered as keywords around which all the configuration, execution and results are organized. Thematic categories are also used to classify the characteristics (expected values) of the datasets required for the application context (TDS_i) thus allowing to organize the expected quality matrix, and to assess the mandatory and recommended metadata/quality elements for specific metadata profile(s). An example of thematic categories that can be used to categorize the targeted datasets defined by the user is the list that integrates the INSPIRE directive (Tab. 3-6).

Tab. 4-2. Description of metadata/quality elements for metadata quality evaluation.

Metadata/ Quality element(s)	Thematic category applicability	Verification method	Element from metadata profile
Unique identifier	all Thematic categories	All elements filled	Unique resource identifier
Abstract	all Thematic categories	All elements filled	Resource abstract
Lineage	all Thematic categories	All elements filled	Lineage
Limitations on public access	all Thematic categories	at least one element filled	Conditions applying to access and use
DQ: Conceptual Consistency	Thematic category: Species distribution (SD)	All elements filled	Conceptual Consistency with defined 'DQ_EvaluationMethodType', 'nameOfMeasure', 'evaluationMethodDescription', DQ_QuantitativeResult'
DQ: Domain Consistency	Thematic category: Species distribution (SD)	All element filled	Domain Consistency with defined 'DQ_EvaluationMethodType', 'nameOfMeasure', 'evaluationMethodDescription', DQ_QuantitativeResult' and 'result'

4.2 ThemisE platform functionalities specification and user´s requirements

Considering the framework described in the previous section and the main objective of developing a web-based platform to support and facilitate user-oriented quality evaluation routines for a rapid/agile and adequate assessment of the internal and external quality of data based on (spatial) metadata quality evaluation, the *ThemisE* platform design considered the following functional requirements (Fig. 4-2): (i) user-oriented data quality evaluation routines to support metadata oriented search; (ii) customization of the quality indicators and the metadata profile to be used in the evaluation process; (iii) access to metadata catalogues based on communication protocols/standards services with other platforms; (iv) a user interface to allow the definition of expected values for quality indicators for different applications contexts, and, (v) the selection of metadata catalogues and the visualization of reports resulting from evaluation process.

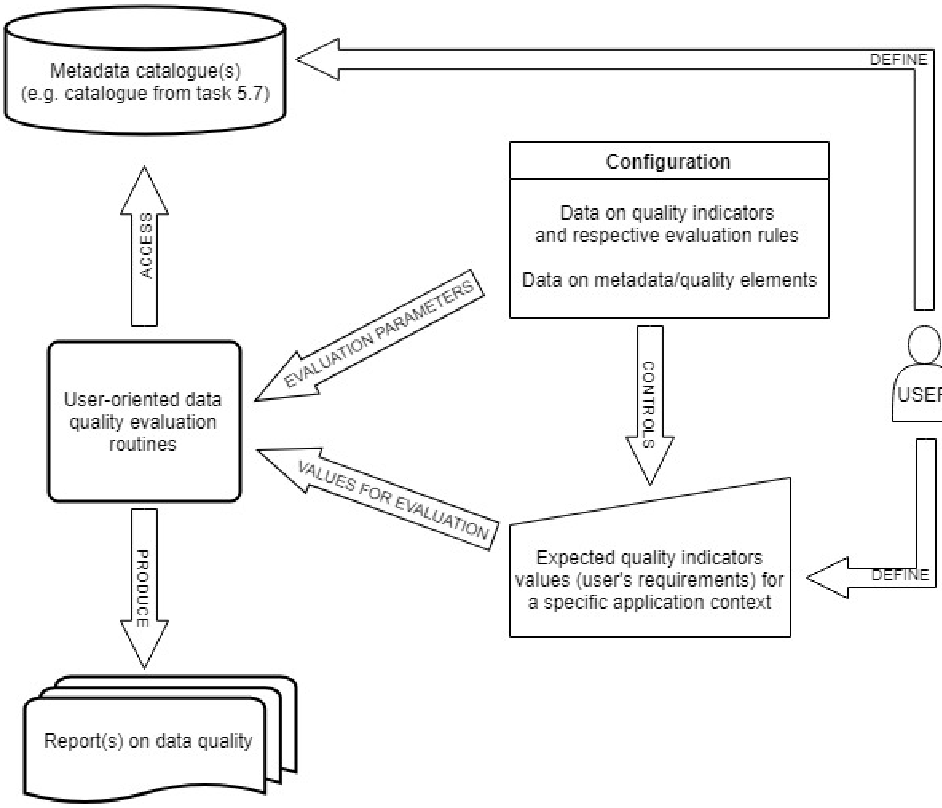


Fig. 4-2. Global functionalities of *ThemisE* platform.

Thus, the development of the Web platform included: (i) the analysis, definition and characterization of the users' requirements in terms of functionalities and uses, considering that these points are critical for the development and implementation of the Web platform, in order to facilitate users' adoption of the platform and its effective use; (ii) the conceptualization of a customizable platform to be configurable to different contexts of quality evaluation; (iii) the development of a generic connection to metadata catalogues; and (iv) the design, implementation and testing of a Web graphical interface associated to the platform functionalities. The development of the platform also took into account: (i) consulting and analysing projects with similar objectives and technological solutions; (ii) analysing metadata profiles according to the specific legal and normative framework; (iii) thoroughly discussing, within the technical team in charge of Task 5.5 (in cooperation with Task 5.3 and 5.7), all issues regarding requirements; (iv) developing a highly configurable platform to be easily adapted to diverse quality evaluation requirements; as well as (v) considering and testing different technological solutions for the platform implementation in order to guarantee an easy integration with the database framework to be developed in task 5.7 and with the service-based platform (the ECOPOTENTIAL Virtual Laboratory Platform), especially on how to get the metadata and how to create data quality routines.

The *ThemisE* platform is based on a user-oriented Web interface that allows users to systematize and configure quality indicators and metadata profiles, define quality requisites, select metadata catalogues and run evaluation routines. Considering the functionalities to be implemented, two types of users could be identified: (i) a standard user which will use the metadata quality evaluation module by defining quality requisitions and running evaluations; and (ii) an advanced user (administrator) that has the role to configure the platform regarding metadata profiles and quality indicators to be used in evaluation processes.

Based on these two user profiles, a set of main requirements were determined for all users to facilitate user oriented quality evaluation, namely the need for tools and interfaces: (i) to define and configure quality indicators for their own applications contexts, (ii) to select catalogue services to be used as metadata sources and, (iii) performing external quality evaluation to search and discover the datasets available that meet the expected selection/evaluation criteria. Beyond these three standard and overarching requirements, the advanced user profile will have the possibility to configure the metadata profile to be used in metadata quality evaluation and the quality indicators rules that can be used in external quality evaluation.

According to the defined requirements, the *ThemisE* platform provides a set of functionalities to allow standard users to perform actions necessary for user oriented quality evaluation (Fig. 4-3), namely: (i) define each quality indicator by selecting a thematic category and configure quality indicators values (allowing to set options such as critical factors, filter criteria and configure expected quality values); (ii) configure and select the metadata catalogues (CSW) to be used in the evaluation process; (iii) manage requisition data by saving, downloading, deleting or uploading the expected values for quality indicators; (iv) run the evaluation process based on defined expected values and selected catalogues; and finally, (v) visualize the results of the evaluation process using different mode views (e.g., tree view, tabular, …).

Fig. 4-3. Standard User use case[37].

With regard to the advanced user, a set of actions has been added to allow platform configuration to be customizable to different application contexts. The configuration process (Fig. 4-4) includes options to: (i)

[37] Use case diagram is a representation of a user's interaction with the system. The round balloons represent the actions the user can perform, balloons with the tag <<include>> are actions that must be performed by the user to accomplish the action connected to that action. The <<extend>> tag means the action may or may not be performed to accomplish the connected action.

define thematic categories that will be used to group expected quality indicators; (ii) define the quality indicators that can be used by the metadata quality evaluation module, including metadata element and rules; (iii) configure default values for controlled lists used in the user interface; and (iv) specify the metadata profile structure and required elements to perform metadata quality evaluation. Note that this configuration process should be executed by a user with advanced knowledge to guarantee the correct configuration of the platform for the intended application context. At this stage, this is accomplished by filing a JavaScript structure (using JSON notation) directly in a configuration file. In the future, a graphical user interface could be developed in order to enable this configuration process by a wider group of users.

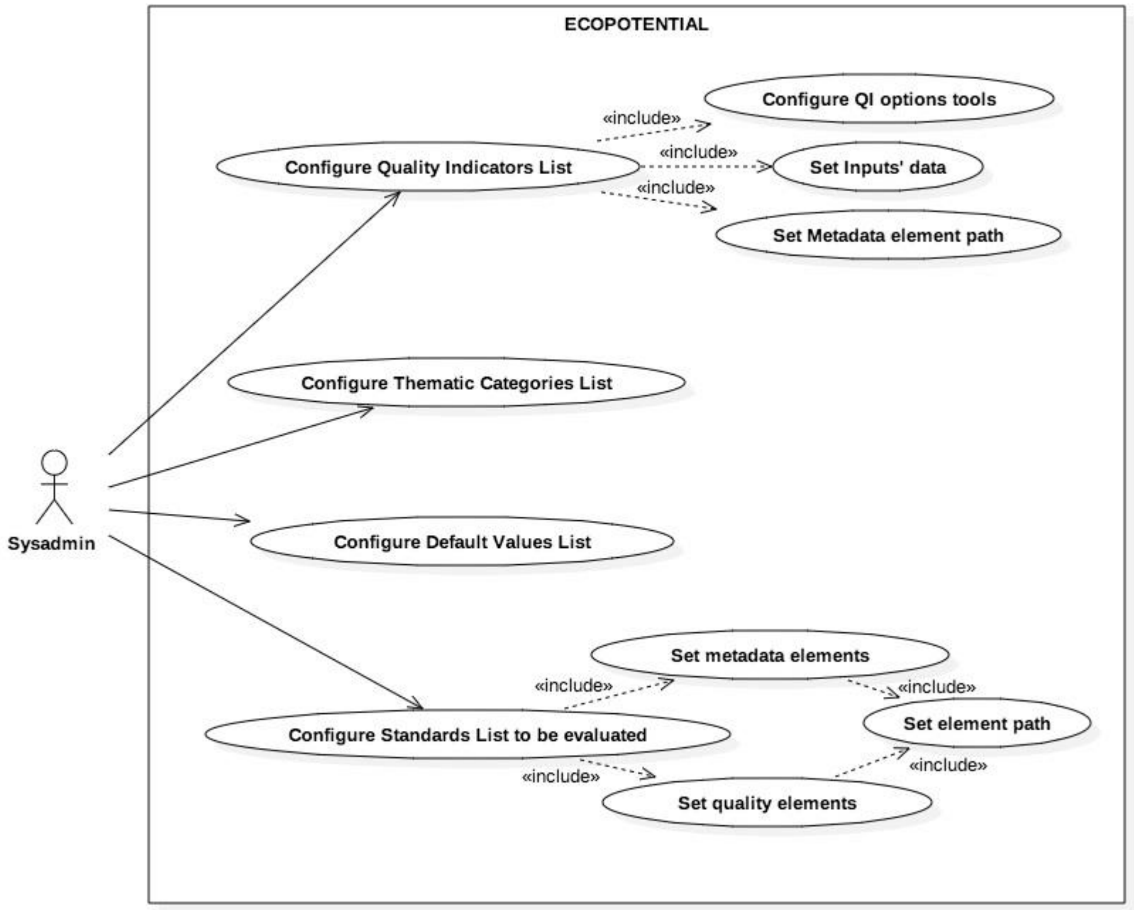


Fig. 4-4. Advanced user (administrator) use case.

Another fundamental component of the ThemisE platform concerns to the services that access to metadata catalogues in order to implement an interoperable solution applicable to diverse contexts and domains of application within and outside the consortium. Thus, CSW OGC standards [Web Catalogue Service] were selected as the core communication service in order to guarantee the use of a communication standard easily recognized by other catalogue services available worldwide, but also to allow an easy integration with metadata catalogues to develop and make available under ECOPOTENTIAL Virtual Laboratory Platform. Regarding the standard for communication with a catalogue of geospatial records through CSW services, the standard(s) used in the implementation of the platform follows the metadata implementing rules of the ISO 19115 “Geographic Information - Metadata” (EC, 2010b) and is encoded according to the implementation schema ISO 19139 “Geographic information - Metadata - XML schema implementation” (EC, 2017; ISO, 2007).

Note that one of the catalogues already configured/tested in ThemisE platform, within the ECOPOTENTIAL project, is a metadata catalogue for in-situ metadata that has been developed in task 5.7. This catalogue follows the CSW standards and supports different metadata profiles, such as ISO 19115 and ISO 19139. It is linked to the ECOPOTENTIAL Virtual Laboratory platform and will be utilizable to support the evaluation of the metadata providing the metadata quality elements described in previous sections. Using the mentioned standards, the ThemisE platform is able to send queries (e.g. GetCapabilities, GetRecords, DescribeRecord, GetRecordById, Harvest and Transaction) to this metadata catalogue in order to identify the datasets that fulfill users´ needs.

4.3 ThemisE platform logical and technological architecture

The ThemisE platform has been developed as a multi-module to facilitate the metadata quality evaluation among project Partners, using real-time data access to different metadata catalogues in order to enable simple and practical assessment of the internal and external quality of spatial data based on metadata and users' needs or expectations. To fulfil the objectives defined for the ThemisE platform, the following core development options were considered: (i) the platform must provide the user-oriented quality evaluation functionalities through the World Wide Web; (ii) the implementation should preferably use Free Open Source Software to provide a low-cost development and maintenance solution; (iii) the platform should be as flexible as possible to be compliant to different uses, allowing to be adapted to different metadata profiles and to integrate a variety of different quality indicators; (iv) the system must be developed to facilitate the interoperability and integration with other internal and external platform, such as catalogue services to allow metadata searching and retrieving and information systems to manage users and respective requisition data for quality evaluation; and (v) the platform must have a graphical user interface that provide a simple, easy and comprehensive view for the definition of user's expected values (metadata requisitions for each dataset) and visualization of the evaluation's result, since some of the platform's users may have no prior experience on metadata.

Considering the defined development options and, in particular, the intent to implement a modular based platform to facilitate the future maintenance, integration with other systems and development of new functionalities, the metadata quality evaluation platform has been implemented as a standalone client application and structured as independent communicating elements to enable straightforward new implementations to integrate with other information systems. Regarding the application development process, the platform was implemented following an object-oriented architecture composed by the integration of different selected, adapted or developed components. This approach permits the implementation of a flexible platform allowing developers to easily upgrade the platform in the future.

From a general point of view, Fig. 4-5 represents an abstract view of the functional architecture of the developed platform. The central element of the platform is the module (Web application on the client-side) that controls the graphical user interface and evaluation routines whose behaviour will be defined through a configuration file to enable different evaluation contexts and quality indicators. This application will be able to communicate with metadata catalogue's servers on the Internet through standard OGC Catalogue Service for the Web to discover and get metadata records to be evaluated based on user's preferences. The application includes also the tools to upload and download files with the definition of quality indicators user´s need (expected values) for different thematic categories to be used in evaluation process.

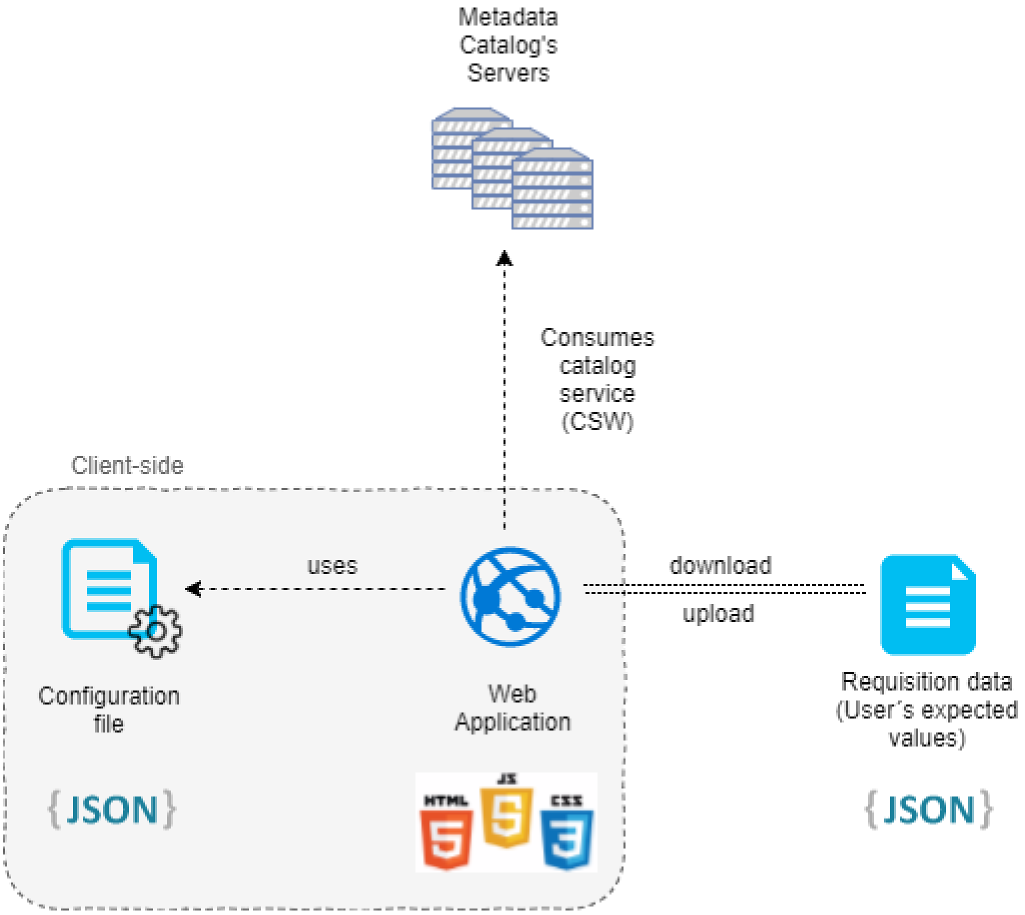


Fig. 4-5. ThemisE platform functional architecture.

From the point of view of the implementation strategy, the metadata quality evaluation platform is based on a structure with three main modules (Fig. 4-6): (i) configuration module; (ii) requisition module; (iii) evaluation routines and a result presentation module. The configuration module allows the advanced user to configure the whole platform's behaviour, including the available quality indicators that may be used in the external evaluation process and the metadata profile to be used for metadata quality evaluation. The requisition module is intended to be used by standard users in order to describe the user´s needs within the framework of its application, being necessary to introduce a set of user-defined expected values for specific quality indicators, to select critical indicators, to define filters for metadata retrieving and to select the metadata catalogues to be used in the evaluation. The evaluation routines and the result presentation module is responsible for the execution of all evaluation routines with no interaction from the user, to produce a result with the overall match of available data and user's needs that can be visualized in multiple view modes.

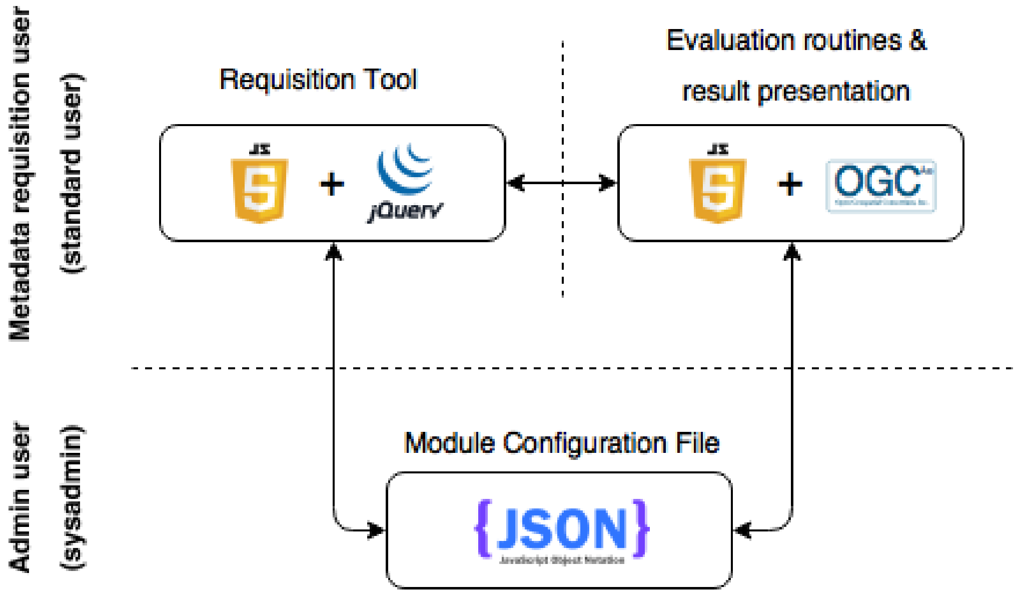


Fig. 4-6. General modules structure of the ThemisE platform.

In order to implement the defined modules, several software technologies and components were analysed in order to select the best options to fulfil the functionalities defined for the metadata quality evaluation module. For the analysis and selection process of the technologies, several factors were considered such as fulfilment of functionalities, knowledge and experience of the development team, compliance with standards, license type, usability, available documentation and community support, usage examples and maturity. Another requirement was to develop a universal platform, with minimum dependencies of other systems.

Considering these requirements, the implementation of the platform as a (standalone) client-side web page was selected, using HTML for the structure and CSS for style and layout. As main programming language to develop the functionalities of all modules, JavaScript (JS) was selected considering that it is one of the core technologies for the production of World Wide Web contents, which allows the platform to be compatible among modern Web browsers. Besides the universal support of JS, the use of this language allows to create an almost self-contained platform executed on the client-side, permitting to have a software solution that is the least possible dependent of web server(s) (excluding the retrieval of metadata records from catalogues or other specific implementations such as login functionalities). With respect to data formats, XML and JSON were selected, where XML is used in metadata documents and communication with catalogue and JSON to define the configuration of the platform.

With respect to data storage regarding the functionalities associated with the management of users' needs or requirements, local storage strategies were implemented at the present stage of the project namely local files and Web storage (within the user's browser). The first one was used to enable the exchange of requisition's data between different researchers allowing to save and load local files, while the second one was used to locally store data between usage sessions using modern browser's local storage capabilities. This allows the user to restore the session when the browser is reset, closed or opened. It is noteworthy that the option to store user's data locally in files does not invalidate the future use of the platform, as the modular strategy used in the system development will make possible to add new component(s) that implement a permanent storage strategy for each user, such as the use of a web service connected to a database management system.

The development of the *ThemisE* platform includes also the use of several JavaScript libraries to create a graphical user interface and to execute processing routines that enables requisition management, evaluation execution and result presentation. JQuery, a JS library, was selected to provide a powerful and easy-to-use Application Programming Interface (API) for HTML manipulation, event handling, animation and Ajax across a variety of browsers. In order to provide a simple and easy-to-use graphical user interface, the platform implementation was based on an extensive use of several graphical libraries with the intent to create a web-based desktop-like application, based on the use of graphical controls such as text fields, combo boxes, tab panels, grids, toolbars, windows, forms, among others. Lastly, OGC Web Services Library for JavaScript developed by the Open Source Geospatial Foundation was also selected to allow the use of Catalogue Services for the Web (CSW) according to the OGC specifications (OGC, 2007), in order to allow the discovery and retrieval of metadata using a standard request-response model of the HTTP protocol between a client and a server using XML.

4.4 ThemisE platform implementation and functionalities

As mentioned, the ThemisE platform should provide the necessary instruments for two main use types: evaluation and configuration. The former should provide tools to define the user's implicit and/or explicit needs or expectations for a given application context and geographic area and to evaluate and view the degree of matching between the characteristics of the data (as detailed by metadata records obtained from catalogue services) and the characteristics required by the user. The latter should provide a procedure to configure the evaluation process to be applied for different applications contexts, namely to define the metadata/quality elements required according to a specific metadata profile, the available thematic categories, the selectable quality indicators that can be used to evaluate each metadata adequacy to the user's needs, and to set the default values lists for specific fields.

According to these requirements, this section describes the ThemisE platform and discusses issues related to: (i) platform implementation; (ii) user's requirements input and management; (iii) evaluation process and results visualization; and (iv) platform configuration.

4.4.1 ThemisE platform implementation

As previously stated, the ThemisE platform provides a simple and easy-to-use graphical user interface to facilitate the interaction by less-experienced users who are unfamiliar with metadata and data quality evaluation concepts. Therefore, the platform implementation was based on the use of several graphical libraries with the intent to conceive a rich web-based application. The human-interaction interface fell back on the integration and customization of components of free and open source software and applications, namely on the combination of several graphical user interface (GUI) controls (or widgets) to enhance the efficiency and ease of use: Bootstrap to build a responsive front-end user interface; JQuery UI, to build an interactive desktop-style user interface, based on the combination of customizable visual widgets; Tabulator to create interactive tables; and Bootstrap Tree View to display hierarchical tree structures. For the implementation of interactive maps, Leaflet library was selected to implement the geo-location manager. Leaflet is an open-source JS library for mobile/web-friendly interactive maps allowing viewing and editing geospatial data, with an extensive range of features available through a variety of plugins.

One of the critical issues of the *ThemisE* platform is related with the definition of user's needs or requirements to be used in the evaluation process. Therefore, in order to implement a simple and user-friendly definition process that copes with possible difficulties of less-experienced users who are unfamiliar with metadata profiles, a matrix-view approach was developed to facilitate the definition of the desired themes and respective quality indicators. The implemented approach was based on the development of a grid layout, which allows defining themes to search by columns and quality indicators by lines. In order to facilitate the definition of simple and more complex quality requisites, a custom editor was implemented which allows criteria to be edited on-line and validated by automated procedures.

Through the use of the ThemisE platform interface, the user can perform a series of actions in order to search and evaluate data sets in terms of data availability and adequacy according to data requirements, that includes the definition of user´s needs, consisting of the specification of the intended themes and their quality indicators, the selection of metadata catalogues to be searched, the execution of the evaluation process and the visualization of the evaluation results (Fig. 4-7). The minimum requisite to perform an evaluation includes the setting of at least one metadata database catalogue and one theme of a thematic category with one quality indicator. For each theme to be considered in the evaluation process, the user must select a thematic category and then define expected values for the different available quality indicators that are pertinent for the context application.

The definition of the expected values has to be done for each relevant quality indicator, considering the specificity of each data set type and the application context. The specification of quality indicators includes the logical combination of values and the option to set it as a critical factor and/or as a filter. Another action refers to the setting of the URL of metadata database catalogues compliant with OGC CSW and the selection and test of catalogues to be used. After completing the previous tasks, the evaluation process can be initiated. At this point, each metadata record is retrieved from selected metadata database catalogues and evaluated against user's defined requirements. Once the evaluation has been completed, the user can select different view modes allowing analysing which datasets are fit or unfit and to what extent they are suitable or unsuitable for a given purpose, as well as, a summary describing the criteria that exhibit potential problems for attaining predefined quality goals.

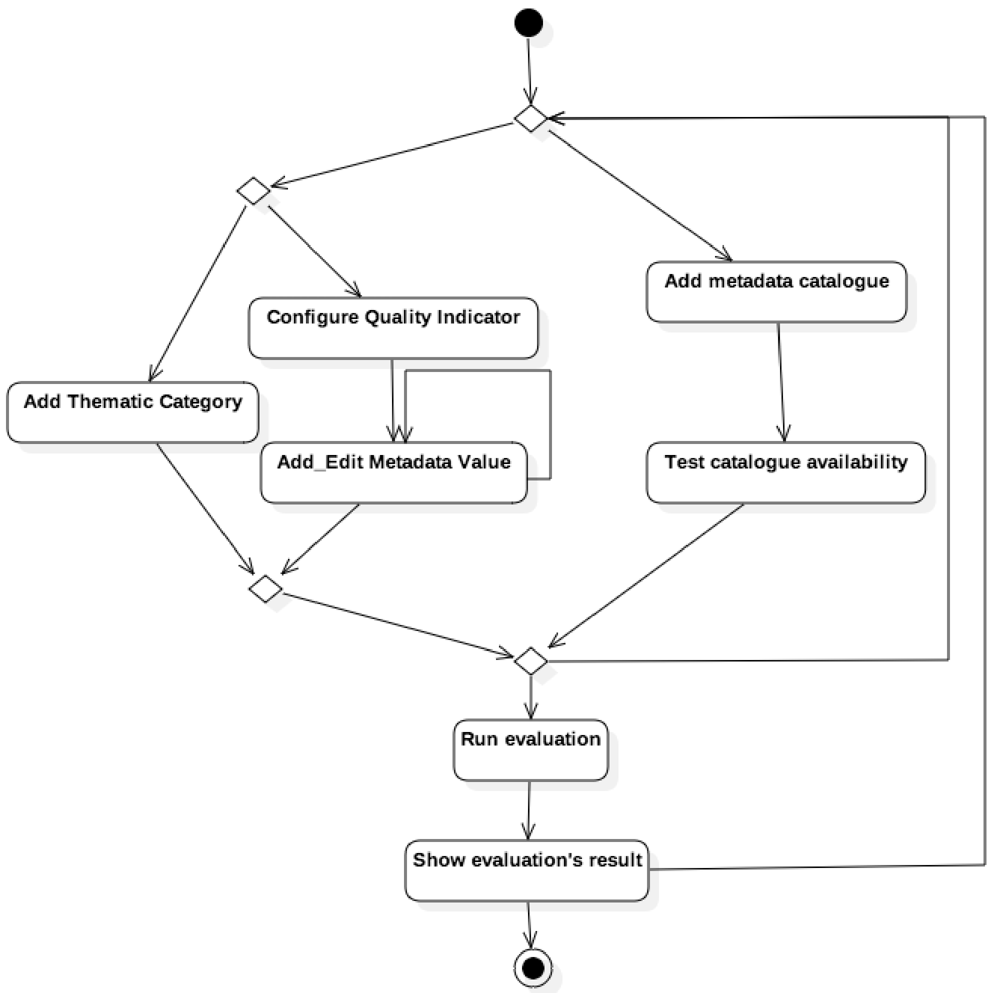


Fig. 4-7. ThemisE platform actions workflow.

The evaluation process is managed by the platform and involves the execution of a sequence of five phases to allow the visualization of final results in different layouts (Fig. 4-8): (1) fetch datasets' metadata from database catalogues; (2) evaluate required metadata values; (3) filter dataset's metadata according to user's filters; (4) evaluate metadata/quality elements according to specific standard(s); (5) visualize evaluation's results.

The first phase involves fetching the dataset's metadata from the different database catalogue's servers configured and selected by the user. In phases 2 and 3, each downloaded metadata is evaluated according to the defined user's requirements, being the metadata elements compared with the respective quality indicator expected values defined for each thematic category in order to determine the level of conformance. In case of the existence of a non-conformity for a quality indicator defined as a filter for a particular thematic category, the dataset is removed from the result list for this thematic category. The fourth phase comprises the assessment of each downloaded metadata with regards to the fulfilment of the required metadata/quality elements according to the standards configured by the advanced user. The last phase includes visualization tools to present the result in different view modes in order to facilitate the interpretation and identification of the more adequate datasets for the case in study. Fig. 4-9 presents the steps involved in the first four phases in the form of pseudo code.

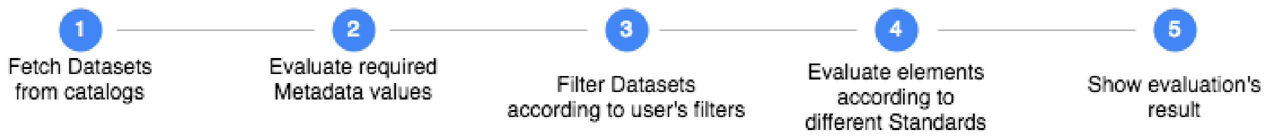


Fig. 4-8. ThemisE platform evaluation process phases.

```
Step 1
for each SRC in Dataset's Source:
  fetch all datasets from SRC


Step 2, 3
for each TC in Thematic Category List:
  for each SRC in Dataset's Source:
    for each DT in SRC's datasets:
      for each QI in TC's Quality Indicators:
          compare requested values (QI) with dataset values (DT)
        if QI is used as filter and QI is not in conformity:
          exclude DT from QI's datasets


Step 4
for each TC in Thematic Category List:
  for each DT in TC's datasets:
    for each STN in Standard in configuration file:
      for each ELEM in STN's elements:
        check if ELEM is in DT structure
```

Fig. 4-9. ThemisE platform evaluation process pseudo code.

Regarding the software development, it should be pointed out that an agile approach has been used in order to facilitate the integration of new requirements and continuous improvement/update of functionalities, as the solution evolved as a result of the collaborative effort of the team. Thus, the platform development included several testing phases executed by users, not involved in the software development, that enabled the identification of limitations and inadequate user interface options, for example, the necessity to implement more than one communication mode with metadata catalogues, the selection of a more user-friendly date picker, the visualization of dataset names recorded in metadata files, among other aspects.

Finally, considering that European-funded H2020 projects such as ECOPOTENTIAL should have as a main concern the sharing of knowledge, the full source code of the actual version of ThemisE platform is licensed under the GNU General Public v3 License and is available in a GitHub repository, in order to promote internal and external collaboration (https://github.com/pmrcastro/ECOPOTENTIAL-WP5.5-ThemisE[38]).

4.4.2 User's requirements input and management

From a general perspective, the graphical interface of the platform consists of a matrix presented as a grid page with individual toolbars in each grid cell to control their contents and a global toolbar at the left side with global functionalities (Fig. 4-10). This layout of the ThemisE platform provides the necessary tools to set

[38] Temporary personal repository – to be included in a ECOPOTENTIAL project repository.

the user's requirements, to search for the datasets necessary for a specific application context in terms of quality indicators, and to define and select the sources of metadata database catalogues. In this sense, two interface sections can be identified: the 'quality definition grid' section and the 'catalogues sources setting' section.

The quality definition grid section provides functionalities that focus on the specification of expected values for the set of relevant quality indicators (QI), allowing combining multiple match operations, flag important QI as critical factors and filter datasets. The grid page is organized in columns that corresponds to datasets to search/identify and in lines relatives to the available configured quality indicators (Fig. 4-10). Each grid cell offers the tools for the definition of expected values for a particular quality indicator (line) of a specific targeted dataset (column). Additionally, each cell displays a summary text with information of the number of specified values. This option to display the information in a grid arrangement allows users to have a global reading and perception of the relevant characteristics of the necessary data for the application context under analysis. In addition, this matrix layout facilitates the definition and checking of the expected spatial data quality values.

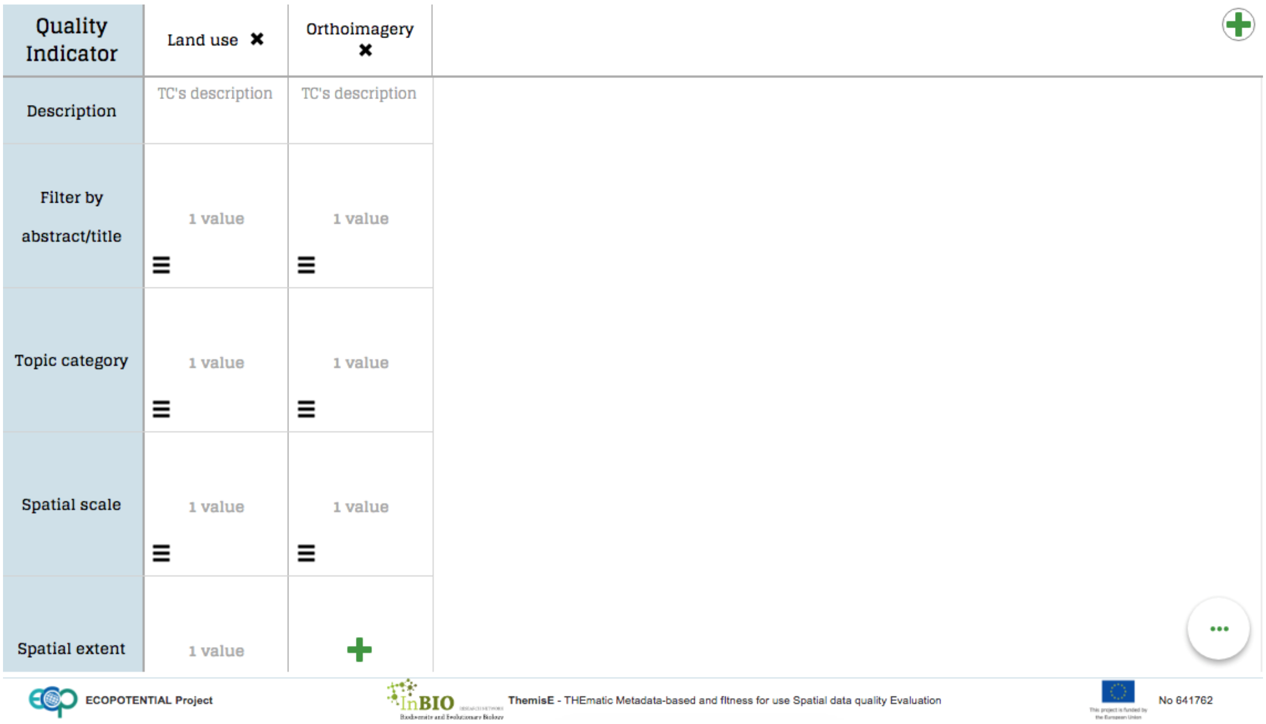


Fig. 4-10. Expected quality definition layout page.

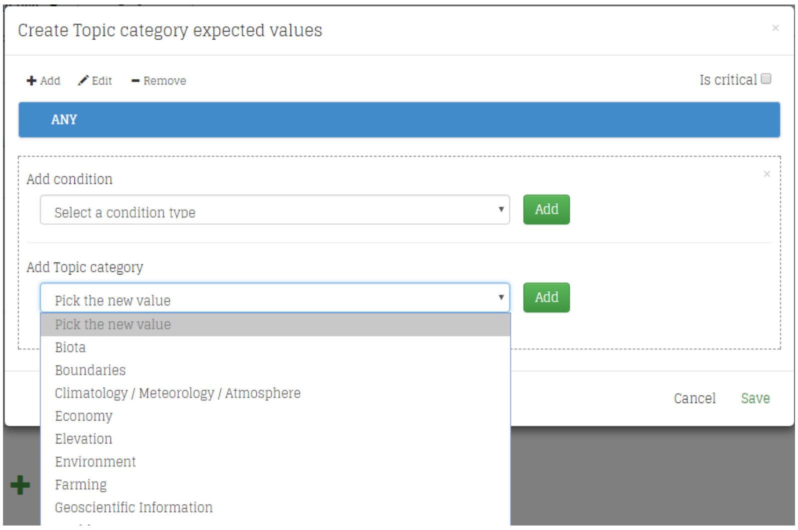
The filling of the quality grid page begins by adding a new column and selecting a topic category that identifies a dataset targeted for the context application. Then the user can specify the expected data quality values for the applicable quality indicators of the targeted dataset. This process can be repeated as many times as necessary to define all the datasets quality characteristics needed. It is important to note that the adequate accomplishment of the quality matrix filling strongly depends on the user's expertise regarding the application context.

Considering the fulfilment of the expected data quality values of available quality indicators (corresponding to the cells of the grid), the accessible functionalities allow the user to add new quality indicator requirements, to edit or remove defined expected values and to copy an existing definition of expected values of a topic category to another topic category, corresponding from the point of view of grid layout, to copy the definition of a cell between columns of the same row (this option is useful for example to copy expected values of spatial extent quality indicators between the different targeted datasets).

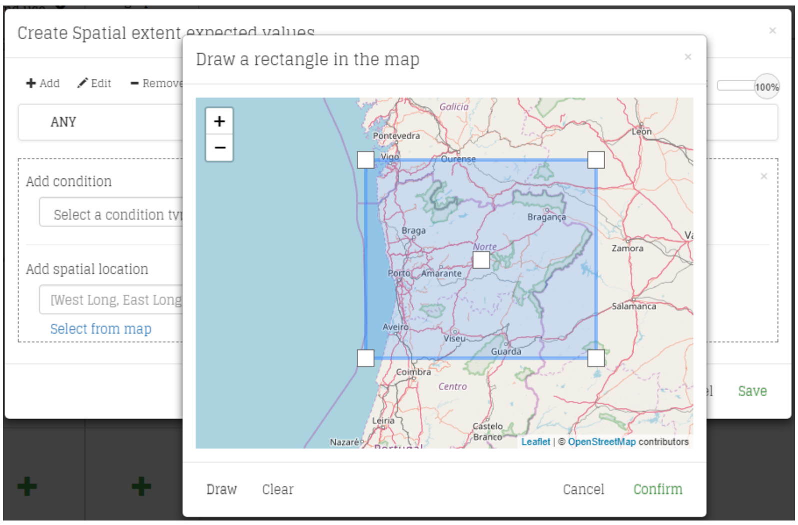
For the specification of each quality indicator requisites, the user has access to an on-line web form (Fig. 4-11) with the appropriate options and graphical user interface component for the entry of expected quality values according to the indicators' data types. The forms gives the user the possibility to manage expected values definition by filling or selecting values using text fields, lists, date fields, but also the ability to specify the behavior (rule) when combining multiple values for a quality indicator to allow the definition of more advanced conditions. The two available options to control multi-value comparisons are: 'ALL' used to specify that all values must be verified and, 'ANY' to indicate that at least one defined value must be matched when comparing expected values with metadata values. For example, Fig. 4-12 presents a multi-value definition for topic category specifying that quality indicator will be conformant if metadata element has classification with a value of ("Inland Waters" or "Oceans") and a value of ("Transportation" or "Utilities/Communications"). Moreover, whenever possible the quality expected values definition form presents a list of predefined values to support automatic fields' completion (Fig. 4-13a) and provides auxiliary widgets for geographical and temporal extent definition (Fig. 4-13b).

Fig. 4-11. General layout for quality expected values specification.

Fig. 4-12. Setting multi-value comparisons.



(a) list of predefined values



(b) geographical extent definition

Fig. 4-13. Auxiliary components to support quality expected values specification.

In addition to specify the expected values for quality indicators, the user has the possibility to select complementary options to have a more oriented evaluation (Fig. 4-14). These options are dependent of each quality indicator, presented on the top-right corner of the form, and can include:

- Option ‘Use as filter’: when this option is checked, all the datasets that do not verify the defined expected values for the quality indicator will be excluded from the final results. Therefore, this option allows to filter (eliminate) datasets that are not relevant in order to facilitate the analysis of results;
- Option ‘Is critical’: used to specify the quality indicator conformity is critical. Unlike the previous option, the datasets that do not conform with this option will be presented in the final results, allowing the user to identify requirements that are problematic to attain;
- Option ‘Cover (%)’: available for quality indicators with an extent definition such as for spatial bounding-boxes and temporal extents. This option has a default value of 100% that can be lowered to specify the percentage of coverage that is required to classify the quality indicator as a valid result.

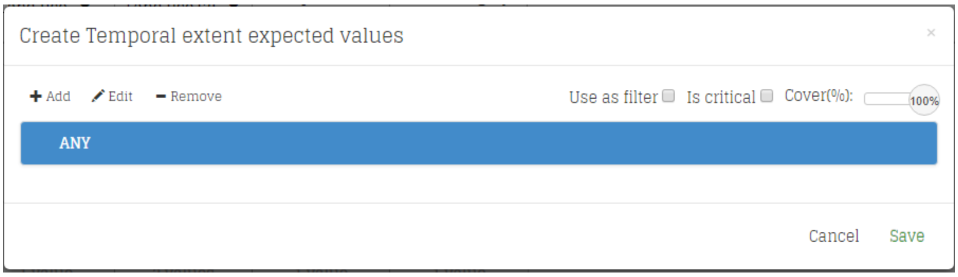


Fig. 4-14. Complementary options for quality expected values specification.

During the definition process of the user’s needs or requirements in terms of targeted datasets and respective quality indicators’ expected values, the user has access to several tools to store the introduced data, namely: (a) to save all configurations done by the user using browser’s local storage capabilities, allowing to restore data automatically when the browser is reset, closed or opened; (b) to clear all data to allow beginning a new process of requirements definition; (c) save all defined data to a local file (download data to user computer) to allow permanent persistence of the information defined for a specific context application; and (d) load a previously saved file through option (c).

In order to be able to perform the evaluation procedure, the user has to setup the metadata database catalogues sources from where the platform will get datasets metadata records to be evaluated based on specified quality indicators (Fig. 4-15). As mentioned previously, the platform has been implemented to communicate to metadata database catalogues through OGC standard catalogue services (CSW). Thus, the setup of a new metadata database sources includes the introduction of the URL to the CSW server endpoint. Additionally, an option was added to allow the user to choose the use of synchronous HTML request for the communication with the catalogue server instead of using an Ajax request since some server do not allow the default option of performing asynchronous requests. As the metadata sources setup can be made for any number of catalogues, the user has the possibility to select (enable) only the sources which are relevant for the application context. When a metadata source is enabled, the platform automatically tries to connect with the catalogue server through the CSW URL to retrieve the total available number of datasets metadata records. This general approach to define datasets metadata sources allows considering the variety of existing catalogues servers on the Internet, such as the in-situ metadata catalogue in development in task 5.7 of WP5.

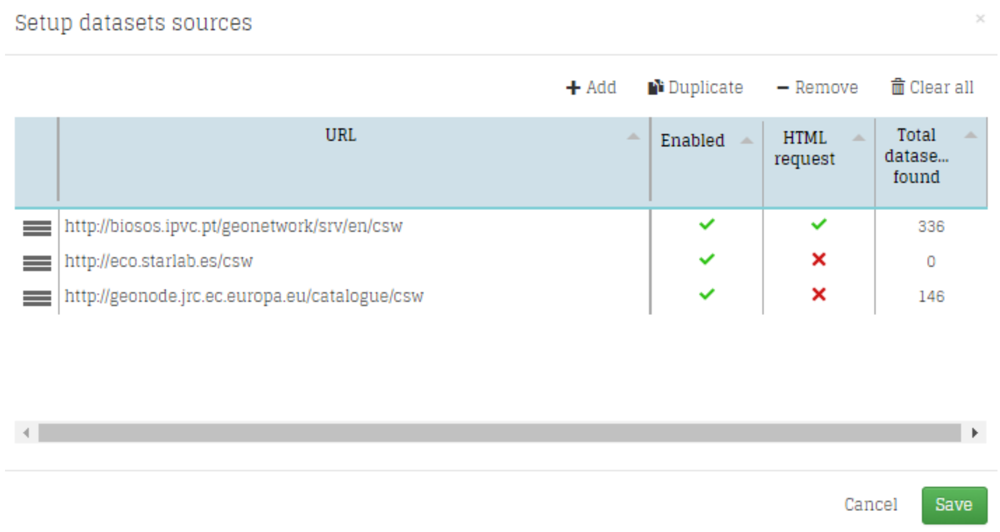


Fig. 4-15. Definition of metadata database catalogues sources.

As a final remark regarding the fulfilment of the quality grid page and considering that the evaluation process can be executed on large metadata catalogues servers, it is noteworthy to emphasize that some functionalities and options have been planned in the *ThemisE* platform to include mechanisms to exclude (filter) datasets from evaluation process to deal with metadata database catalogues with large amount of records. One of these mechanisms is the option 'Use as filter' that can be used in quality indicator specification and previously described. The other tool is the possibility to use a priori filtering mechanism to limit the evaluation process to datasets that match defined values in the abstract or title metadata element. This functionality, identified as 'Filter by abstract/title', is available in the quality indicator list to be easily accessed. Note that this option will only be used as a filter to exclude all datasets that do not verify the defined conditions from the evaluation process and will not be used as a quality indicator.

4.4.3 Evaluation process and results visualization

The evaluation process is fully automatic and transparent for the user. The execution initiates by selecting the appropriate option accessible on the left popup toolbar. During this process, all metadata records of selected catalogues servers are evaluated against the user's needs in terms of datasets, which have been specified through the filling of expected values in the grid page. It should be noted that the evaluation process could be a time-consuming task when using large metadata catalogue server(s), where execution time is largely dependent on the size and access speed to the metadata catalogue server(s), but also Internet connection bandwidth.

As soon as the evaluation is completed, a window with the evaluation results is presented to the user. The results can be analysed by the user from different perspectives selecting one of three alternative visualization modes: (i) summary mode, (ii) statistics mode or (iii) tree view mode.

The summary view mode (Fig. 4-16) presents a list of all targeted datasets (defined as column in the grid layout) with a summary of the fitness for use evaluation for all the datasets obtained from metadata catalogues servers, which comply with the defined filters. The list summary is divided in three groups: (i) fit datasets that includes datasets which are in conformity for all expected values of quality indicators (match of 100%); (ii) partially fit datasets that incorporates datasets which are not in conformity with all quality indicators (at least one non-conformity detected) but are conformant for all quality indicators defined as critical factors; and (iii) unfit datasets which as datasets not in conformity with at least one quality indicator which was considered a critical factor.

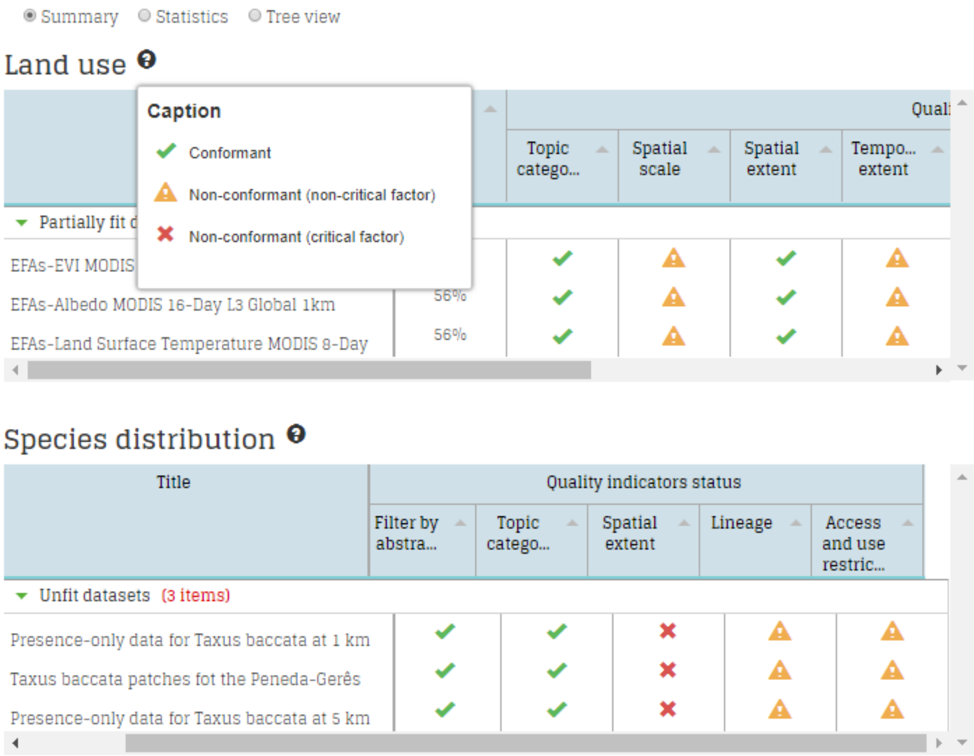
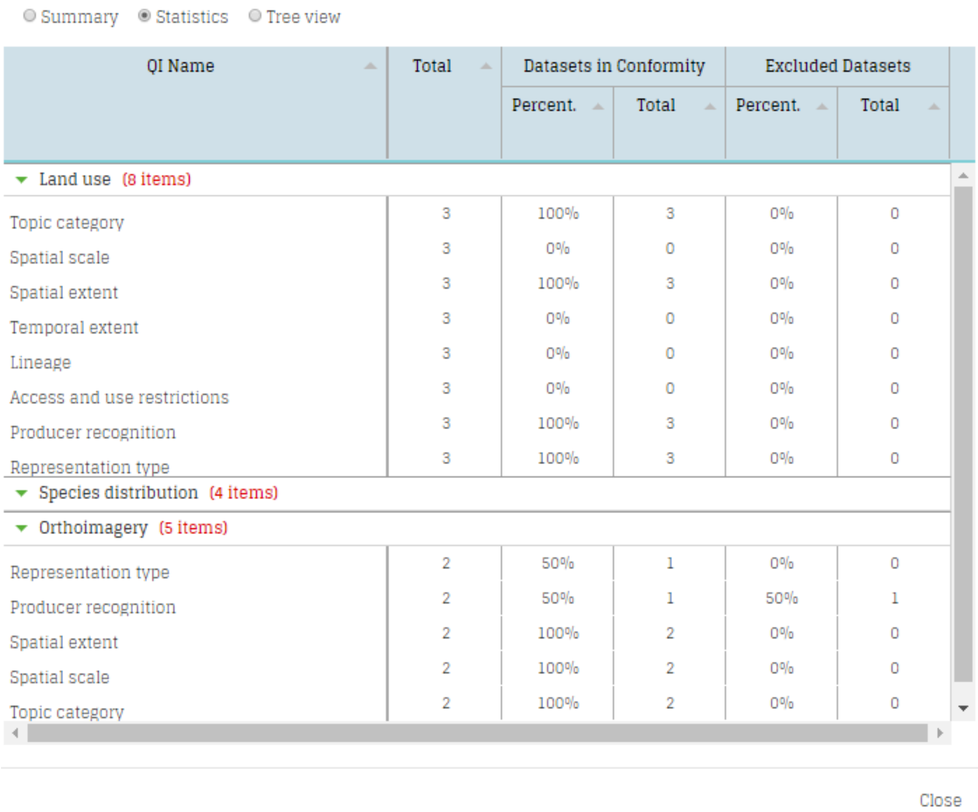


Fig. 4-16. Results summary view.

On the other hand, the statistics view mode (Fig. 4-17) aims to present calculated statistics values about the results in order to facilitate the identification of the quality indicators that have the most influence on the success or failure of datasets availability for the application context. The statistics view presents three values for each targeted dataset and respective to each specified quality indicator: (i) the total number of evaluated metadata of datasets filtered from database catalogue servers; (ii) the total number (and percentage) of datasets (metadata) which are conformant with the quality indicator of the corresponding line; as well as (iii) the total number (and percentage) of excluded datasets for not being in conformity for the specified quality indicator specified as a critical factor.



Summary Statistics Tree view

QI Name	Total	Datasets in Conformity		Excluded Datasets	
		Percent.	Total	Percent.	Total
Land use (8 items)					
Topic category	3	100%	3	0%	0
Spatial scale	3	0%	0	0%	0
Spatial extent	3	100%	3	0%	0
Temporal extent	3	0%	0	0%	0
Lineage	3	0%	0	0%	0
Access and use restrictions	3	0%	0	0%	0
Producer recognition	3	100%	3	0%	0
Representation type	3	100%	3	0%	0
Species distribution (4 items)					
Orthoimagery (5 items)					
Representation type	2	50%	1	0%	0
Producer recognition	2	50%	1	50%	1
Spatial extent	2	100%	2	0%	0
Spatial scale	2	100%	2	0%	0
Topic category	2	100%	2	0%	0

Close

Fig. 4-17. Results statistics view.

The last alternative to visualize results is the tree view mode (Fig. 4-18). This mode aims to allow the user to access all details of the evaluation execution through the use of a tree view which can be expanded to have a more comprehensive view of the evaluation's results. The tree view is structured in the following levels: (i) the first level divides the results by the targeted datasets (thematic category) configured by the user (corresponding to the grid's columns); (ii) level two allows to view results divided by metadata database catalogues selected by the user as datasets sources to be evaluated, being presented the identification of the CSW URL and the total of datasets evaluated; (iii) the third level contains a list of all the datasets found in the catalogue of the previous tree level, identified by the dataset title and two values: the global fitness for use value for the defined quality indicators and the global fitness value for compliance with a selected metadata standard profile; (iv) the fourth level is divided in two groups: (iv-a) a group that is referent to the requested metadata evaluation that allows to consult the conformity and critical factor of each quality indicator resulting of the comparison of user defined expected value(s) and metadata element value(s) of dataset identified on the third tree level; and (iv-b) a group that presents the compliance of each metadata element of the dataset (identified on the third level) with respect to the metadata/quality elements for each configured metadata standard profiles (the configuration process of the standards to be analysed are described in next section).

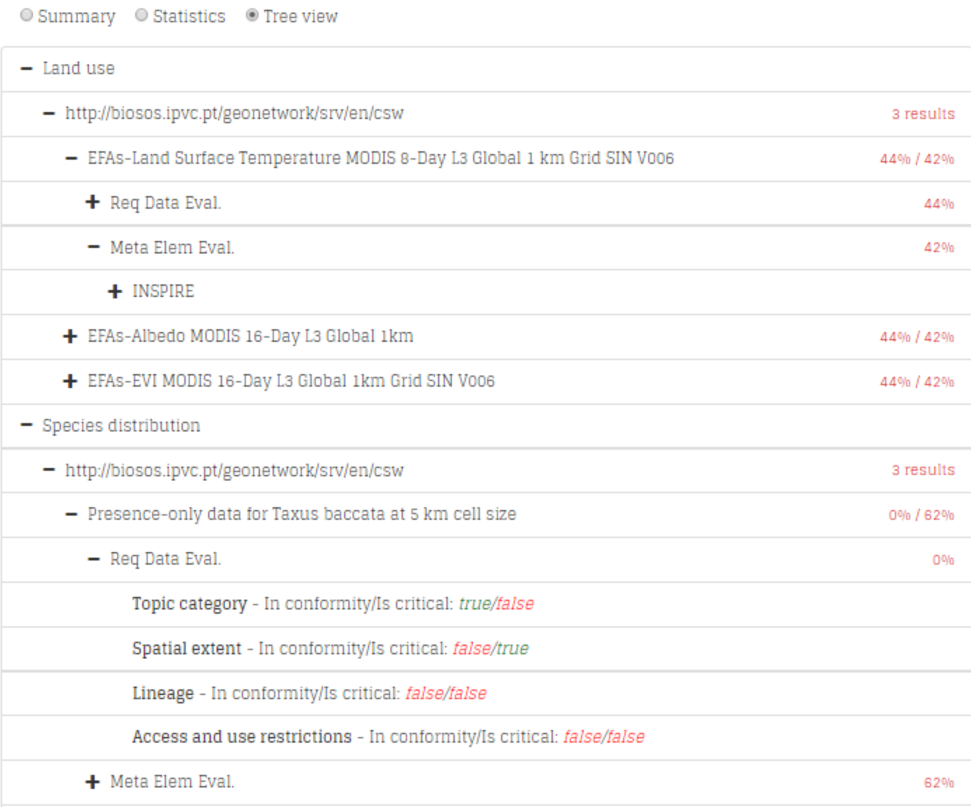


Fig. 4-18. Results tree view.

Besides the results' visualization modes described, the platform allows to individually view the metadata contents of each evaluated dataset obtained from catalogue servers by clicking on the dataset title on the summary and tree view modes (Fig. 4-19).

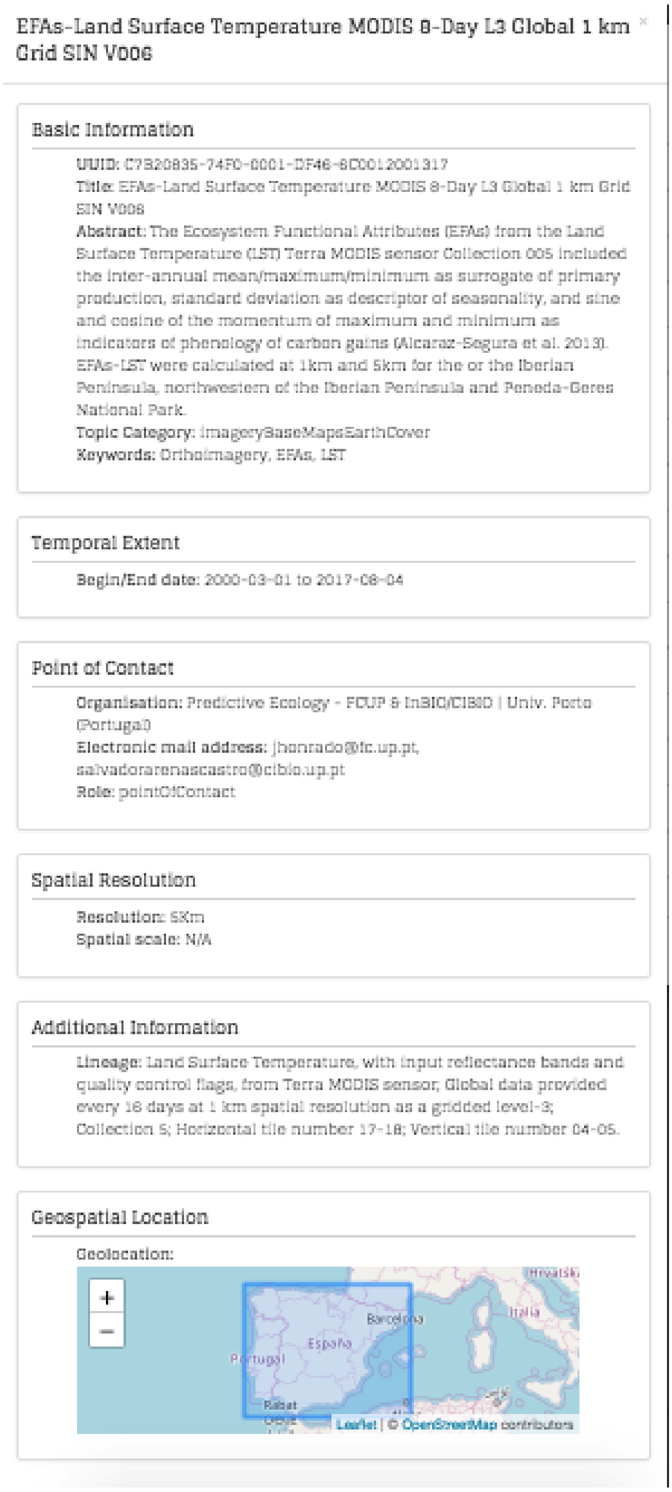


Fig. 4-19. Metadata contents view of evaluated datasets

4.4.4 ThemisE platform configuration

Within the scope of the ThemisE platform development, one of the concerns was to implement a highly customizable platform to be adjustable to a large variety of application contexts but also to be easily adaptable to a large set of external quality evaluation contexts, namely, to be flexible to a variety of metadata standards profiles and quality indicators.

At this stage of the platform development, a critical issue for the platform customization is the need to implement a solution as universal and flexible as possible to allow developers to easily upgrade the platform in the future. Considering this requisite, the implementation options were based on the specification of a set of parameters in a specific format in a predefined file of the platform, since this choice allows having a highly adaptable and configurable solution. The chosen format for the present software was JavaScript Object Notation (JSON), as it is a lightweight data-interchange format with the characteristics to be easy to read and write for humans. This approach has the disadvantage that the platform customization needs the intervention of a more advanced user with some knowledge in JSON notation and interpretation. Nevertheless, this limitation can be easily overcome through the development of a graphical interface to offer a more user-friendly customization process.

The configuration file content allows customizing the following properties of a global item for the ThemisE platform: (1) thematic categories, (2) predefined values of specific input fields; (3) metadata/quality elements requisites for standard metadata profile(s) to be used in evaluation; (4) quality indicators available for expected values specification to be used in evaluation process; and (5) visual definition of metadata elements to be used in datasets metadata contents view. Each property customization is achieved through the definition of JSON items with a predefined structure.

The first and second item allows the configuration of the available values for a specific application context. The former specifies the thematic categories that can be used in the addition of targeted datasets and in metadata/quality elements requisites definition for standard metadata profiles. The latter allows the configuration of lists of suggested/possible values to use in some specific input fields of the platform forms, such as values for topic categories or spatial reference systems.

The third item allows the configuration of the necessary parameters to evaluate the fulfilment of the metadata/quality elements of evaluated datasets with respect to a specific standard profile. For each standard profile, a JSON item must be configured with the metadata/quality elements with respective XML path, common and specific to the thematic categories configured previously.

The configuration of quality indicators is made through the filling of a list of items, one for each quality indicator, that includes the following global properties: the identification of the quality indicator, the data type (free text, predefined values from a list, scale/range values, time interval, bounding box), the metadata elements to be evaluated, the options to include (filter, critical factor and cover percentage) and available functionalities. Note that the implementation strategy used for the platform development also includes the possibility to configure new data types for quality indicators.

4.5 Spatial (meta)data quality management and collaborative e-science

Spatial (meta)data quality evaluation should include (meta)data quality management ECOPOTENTIAL routines and procedures aiming to promote collaborative e-science at ECOP project and earth digital information infrastructures initiatives. (Meta)data quality management (ISO 19158) and data/information system quality management/governance (Information Security Management System; ISO 27001:2005) as well as organizational total quality management (ISO9000 series standards) should involve establishing data quality management technical guidance, data quality control and assurance procedures and practices in order to improve and maintain (meta)data time series consultable, available and accessible that´s helping the users to: (i) be consistent in the way tasks are performed; (ii) reduce the chance of expensive mistakes;

(iii) use time and resources more efficiently; (iv) monitor and improve user satisfaction; (v) identify new service/business opportunities; and (vi) improve public perception of the resulting products (Van Oort and Bregt, 2005).

Implementation of spatial data quality evaluation and system quality control and assurance protocols should be coordinated with similar existing (and proposed) systems and processes within the project. The advantages in integrate all quality assurance project activities within **a single project quality assurance framework** (Victorian Spatial Council, 2009): (i) having an integrated assessment of all quality indicators; (ii) improving data interoperability by defining comparison terms consistently throughout the project; and (iii) creating routines for (meta)data quality evaluation across work-packages/tasks considering all datasets, users and potential uses.

*S*patial data quality management should integrate future ECOPOTENTIAL activities and should be integrated into the **project Data Management Plan (WP10),** in order to allow a more effective implementation of the data quality evaluation and management beyond the project scope and time frame. Project information system that considers, not only functions related to data repository and organization, but also the standardized fulfilling of metadata and the implementation of data quality procedures (internal and external quality). The project could contribute to user (researchers and end-users. e.g. PA managers) data quality capacity-building, by promoting qualification interests and workshop/courses opportunities, and gains with external professional certification (e.g. according to ISO 19122:2004 "Qualification and certification of personnel").

The definition of **data quality policies and guidelines** should be translated in published documents but also be included in proposals and guidelines about pre-existing data property and custodianship (e.g. awareness, access, pricing, licensing, privacy, and confidentiality). More specifically, this should consider: (i) typifying user's access to the identified datasets and other technologies, as well as continuing to explore joint acquisitions and licensing for each partner and site; (ii) defining field (meta)data collection processes and protocols, image processing methods and standards, and modelling frameworks; (iii) establishing property and exploration guidelines for new products delivered within ECOPOTENTIAL (considering the project public funding and the collective nature of tasks and products); and (iv) defining policies and practices of spatial data publication and sharing among partners and users, as well as with information systems external to the project.

Considering these orientations, **data quality monitoring and reporting** should imply: (i) the definition and specification of spatial data models (INSPIRE Thematic Working Groups, 2011) and formats, as well as the development of internal data communication procedures; and (ii) the establishment of quality evaluation procedures across the groups/ECOPOTENTIAL WP project in order to collect and report on (meta)data quality indicators as wells, contributes to implement and maintain (spatial) metadata catalogues.

The ThemisE (spatial data quality evaluation platform) (as an internal/end-product of ECOPOTENTIAL integrate with WP10 Virtual laboratory) and associated concepts, procedures and practices: (i) define common languages, standards and establish quality assurance as support data reality models, communication and decision processes that's includes and holds individual, groups and organizations; (ii) integrate and facilitates cross public and state level obligations and promotes volunteered, citizen and collaborative science; (iii) facilitates multidisciplinary integration with space as common and integrative dimension; (iv) support (meta)data communication in distributed, dynamic with heterogeneous environment; (v) permits asynchrony quality report and communication between uses/users; and (vi) facilitate organization, human and technological systems governance/ management.

Spatial data quality evaluation and management contributes on propose and implement a vision and effective development of Digital Earth. This implies (meta)data standards production, management, sharing as well as, define access/availability/security, property and individual rights management relates to spatial (time) datacube (series) production (structural approach) and multi-purpose digital quality aware metadata repositories/catalogues about reference/multidisciplinary themes (Marine, Geology, Atmospheric, Planetary and Cryospheric science domains) promote data interoperability and data use.

Open physical advanced e-infrastructures with a political and human adequate context which include technological and procedural definition perform Big, Linked and Open data analytics processing techniques from dataset/databases acquired and generated through observations and modelling simulations. Spatial (meta)data access implies interoperability and data sharing which highlight (meta)data repositories, process and geo web services quality, as well (meta)data and geowebservices availability, usability and usefulness challenges centred on user's perspectives.

E-infrastructure promotes e-science including citizen science, reinforce resulting information uses and social impacts as well, promotes the development of support decision systems promoting groups, communities and society involvement and accountability. The OECD (Organisation for Economic Co-Operation and Development) Principles and Guidelines for Access to Research Data from Public Funding consider as benefit of the digital repositories: (i) accessibility (ii) improve public returns of the public initial investment in factual information; (iii) the creation of value chains of innovation; (iv) the enhancement of co-operation value. In fact, improve access and share open data quality encourages: (i) critical think; (ii) diversity of scientific analysis and opinions; (iii) promotes new research, tests of compares alternative hypotheses and methods as well as, new data and scientific education/inclusion. Sharing and open access to publicly funded research data not only helps to maximize the research potential of new digital technologies and networks, but also provides greater returns from the public investment in Science.

5. Conclusions and Outlook

In environmental analysis, modelling and assessment, spatial data documenting different aspects of the environment (e.g. land cover) are frequently rated as factual data. Therefore, metadata including data quality indicators are crucial to assess their fitness for use in different application contexts (Pôças et al., 2014).

5.1 Main contributions and conclusions

The rapidly growing amount of spatial databases produced, published and shared, together with an exponential increase of users and applications of spatial data, raises the need of data quality assessment and management procedures and tools. The ongoing development of geographic information systems (GIS) and spatial data infrastructures (SDI) further highlight the growing importance of, and challenges involved in, implementing spatial data quality assessment concepts and their underlying practices. Knowledge about external quality evaluation concepts, models and tools precedes the definition of (meta)data profiles, the experimentation of evaluation tools and proposals of spatial data quality management. The framework and task results described in the current report are intended to contribute to the scientific advance in the development of concepts, procedures and evaluation tools for spatial data quality, associated with the formulation of new guidelines for quality management. These efforts are crucial for individual and collective capacity building in the context of ecological modelling and monitoring.

Therefore, measuring, assessing, managing and communicating data quality is important throughout the data life cycle. Data quality can be categorized in two broad and complementary approaches: (i) Internal (or normative) quality, related with the intrinsic characteristics of the data as described at the producer level and usually detailed in metadata; and (ii) External quality (user oriented and usefulness quality), which refers to the level of similarity between the characteristics of the data and the user's needs in a specific context of application, i.e. fitness for use. Both approaches have received advances and attention in geographical information science (ISO 19157), with an emphasis so far on the evaluation of internal quality and external quality (fitness for use) associated to Data Control and Data Assurance within a data quality management framework (ISO 19158).

The ECOPOTENTIAL Task 5.5 and the resulting report D5.3 '**Development of data quality evaluation routines**' are focused on external quality evaluation (based on user's requirements). Nevertheless, beforehand internal quality information (data product specification) must be provided in order to enable effective and useful external quality assessment. In the scope of this task, the assessment of external quality of datasets is based on measuring the matching level (fitness for use) between the characteristics of the data, as detailed in metadata entries (often described at producer level), and the characteristics of the data as required by users for a specific application context (quality requirements defined through expected values for predefined quality indicators).

The evaluation of indirect and external data quality of spatial datasets and databases therefore requires complete metadata (organized at standard metadata catalogues) that contain information about the content, data quality, access and use conditions, and other characteristics of datasets. This information can be used for external (meta)quality evaluation but also for knowledge discovery, indexing and searching. Metadata are aimed to facilitate the dialogue between the data user(s) and the data producer inter-alia by providing information (in machine and human readable format) that supports the assessment of the (spatial) data content and quality in relation to their intended use and application. The different metadata profiles analysed and compared (INSPIRE, DEIMS or DEIMS-SDR MD adopted in ECOPOTENTIAL, see section 3.1 and

3.3) in the project context had their limitations with regard to data quality description. Therefore, adaptations and extensions were proposed in order to fulfil relevant data quality standards.

The community metadata profile for datasets proposed by Task 5.3 (see report D5.2, Metadata for pre-existing datasets) focused on the documentation of the dataset content and access constraints in order to enable effective discovery and reuse of data. The profile definition included the documentation of reference data as well as pre-existing in situ data and EO data used in different ECOPOTENTIAL WP. The DEIMS-SDR Dataset Community Metadata Profile is based on extended EML and allows the proper syntactic and semantic mapping to ISO19115/19139 elements supporting the implementation of INSPIRE (ISO GMD[39]), ISO19115 (ISO GMI[40]) and EML[41]. In DEIMS-SDR Dataset Community MD Profile all metadata elements have been mapped to corresponding metadata elements in ISO19115/19139 and EML. The conformance of an ISO 19115 metadata set to the ISO 19115 does not guarantee the conformance to INSPIRE, but the use of guidelines to create INSPIRE metadata ensures that the metadata is not in conflict with ISO 19115.

The use of these metadata profiles and their implementation rules in data catalogues allow the implementation of metadata-based evaluation and quality evaluation routines. However, the number of indicators that can be calculated can be reduced as they depend on the missing or properly completed (either required or optional) metadata elements needed to implement fitness for use quality evaluation metrics.

Examples showing the use of the ThemisE platform (standard INSPIRE) (see section 4) feature the fact that asking about some optional fields would extend the possibility of "quality indicators" to be applied. The current version of ISO 19115-1: 2014 and INSPIRE versus ISO / TS19139: 2007 (Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO / TS19139: 2007 '(Date of revision: 2017-03-02), data quality metadata for ISO 19157 (which are considered optional elements) in catalogues could open up possibilities for the extension of the number and type of quality indicators to be implemented in more complete and robust fitness for use quality evaluation exercises.

An online questionnaire with closed and open questions, disseminated to the ECOPOTENTIAL researchers community, showed:

(i) an interest in knowledge about data quality, including quality theory, concepts, elements (spatial data user´s theoretical knowledge) of several internal ECOPOTENTIAL data providers/data users;

(ii) limited experience and practices (i.e. practical experience) related to (spatial) data quality assessment and management routines (spatial data user´s practical experience) namely involving ISO and other reference standards/procedures; and

(iii) a high awareness, interest and willingness to implement data quality routines (spatial data quality user´s perceived utility).

A general analysis of the respondents' feedback revealed that only a limited number of ECOPOTENTIAL researcher show robust theoretical knowledge and experience in data quality evaluation/management. Nevertheless, as result of the questionnaire a general interest in implementing and foster data quality management practices can be assumed, with special focus on:

(i) development of data quality evaluation tools (e.g. ThemisE platform);

(ii) assertive proposal of data quality management process at dataset, database, task, project,

[39] See example https://data.lter-europe.net/deims/node/8935/iso
[40] See example https://data.lter-europe.net/deims/node/8935/iso19139
[41] See example https://data.lter-europe.net/deims/node/8935/eml

research communities knowledge and collaborative e-science levels;

(iii) as well as contribution to conceptual development and international data quality evaluation standards.

Given the previous considerations, the developed ThemisE platform provides a set of tools to perform external (metadata) quality evaluation to support the quality-driven searching and selection of relevant datasets based on requirements specified by the user, allowing to identify the data (or detect data gaps) necessary for environmental/ecological modelling within the project consortium. Additionally, the ThemisE platform performs a metadata (content) quality evaluation, by analysing the level of compliance between the dataset metadata entries and the metadata/quality elements defined for specific metadata standard profile(s).

The ThemisE platform was implemented as an open modular autonomous web software application with data quality evaluation routines that can be executed independently (requires only the communication with metadata catalogues). It can be easily integrated with other information systems/infrastructures and data quality evaluation processes. Therefore, the platform does not include any user control access or specific user-oriented interface, nor the possibility to manage several user's contexts of application (corresponding to different targeted datasets and expected values for quality indicators) as these functionalities would require to select/define communication protocols to integrate the ThemisE platform with external servers. Nevertheless, this option does not invalidate the usefulness and applicability of the platform, as the adopted software implementation strategy facilitates the future integration with any other information system or external applications due to the modular architecture of the platform.

In addition to user's context management, other future developments or improvements of the ThemisE platform can include:

(i) introducing a methodology to measure the level of compliance for each quality indicator, allowing a more detailed classification of the data;

(ii) using multiple criteria decision methods to cope with the multiple (and possibly conflicting) criteria defined for quality indicators in order to support users in selecting the "best" datasets for the problem under analysis;

(iii) improving the graphical representation/analysis of data quality evaluation results;

(iv) allowing to export the results to different formats (e.g.*.xls; …) that facilitate advanced statistical analysis;

(v) defining additional user profiles accreditation and platform multifaceted user profiles management;

(vi) exploring other data quality sources to text type attributes (e.g. key words and lineage) with emphasis in ontologies analysis,

(vii) and promoting the interoperability of the platform with complementary applications and information systems as well as the integration of data quality evaluation, data quality control and data quality assurance processes.

5.2 Further developments in the ECOPOTENTIAL framework

The concepts and tools developed and described in this report (and more broadly in WP5) should continuously be integrated in the coming ECOPOTENTIAL activities focusing on the harmonization, sharing, and access of (spatial) data by (non)experts. This also includes thematic and non-geographic issues determining the interest and relevance of identifying, assessing, managing and communicating the quality of spatial data throughout the data life cycle. Evaluation, monitoring and quality assurance in spatial (meta)data and spatial data services should consider the complexity of processes and present strategies that take into account scientific, technical, organizational, institutional and economic developments. This challenge implies not only the development of concepts, methods and tools, but also capacity building, collaborative and organizational participation that translates into an effective and efficient operational base.

The results from the work carried out in Task 5.5 raise several questions concerning further developments in the ECOPOTENTIAL framework:

(i) How to promote metadata filling and sharing?
(ii) How to improve metadata quality and completeness?
(iii) How to promote the use of SDQE routines within the ECOPOTENTIAL consortium?
(iv) What are best practices to demonstrate the benefits of metadata and SDQE tools application?
(v) How to define and manage users groups for different application contexts?
(vi) How to continue the improvement and testing of supported metadata profiles as well as the ThemisE platform?
(vii) How to integrate the ThemisE platform in Virtual Laboratory Platform to support data quality management processes?

In order to support the implementation of data quality management guidelines, routines and practices, different activities need to be carried out. This includes capacity building of users (including ECOPOTENTIAL multidisciplinary researchers) with regard to data quality evaluation, the definition of technical guidelines, as well as the maintenance and extension of data quality evaluation tools. In this context, the ThemisE platform can foster the integration and communication between different WPs in ECOPOTENTIAL. In this regard, the integration of the ThemisE service with the ECOPOTENTIAL Virtual Laboratory Platform (WP10) – allowing Partners to use it to perform metadata-based evaluation of data quality under specified application contexts in order to identify available datasets or data gaps based on external quality evaluation concepts – would be a logical next step.

An additional possibility is the extension of the ThemisE functions to exchange with external systems defining user or data profiles with respect to typical data (quality) requirements for model applications. Here, the ThemisE platform is able to support:

(i) the ingestion of metadata from different catalogues using a common metadata profile (link to implemented in Task 5.7);
(ii) the further user-oriented assessment of metadata profiles with regard to relevant quality indicators for data discovery, reuse, and application (link to Task 5.3);
(iii) the assessment of relevant data quality indicators for model applications;
(iv) the testing and evaluation of the ThemisE platform to discover relevant datasets (e.g. modelling) or to identify data gaps.

The application of standard data models (e.g. INSPIRE compliant) and the provision of complete metadata in standardised metadata catalogues/repositories in the long run will support the implementation of data quality evaluation processes and the development of a data quality management technical guidance in ECOPOTENTIAL. This would imply a targeted capacity building effort from consortium users/researchers regarding data quality management.

5.3 Outlook: beyond the ECOPOTENTIAL framework

The Task 5.5 results show links with existing e-science initiatives and collaborative science in the field of biodiversity research and ecosystem services beyond ECOPOTENTIAL. The current version of the ThemisE platform and future extensions are promising advances in data quality evaluation and management (data control and assurance), which is a critical element in e-infrastructure and e-science development. Task 5.5 results intends to contribute to disseminate and vulgarize data quality evaluation principles and tools from SDQE methods/procedures and data quality management process development in open, distributed and collaborative environments including communication and responsibility between researchers, technicians, decision-makers and stakeholders at society and communities contexts.

The current implementation of ThemisE supports data quality management as data quality control and data assurance across ECOPOTENTIAL activities, researchers and users. The adoption of data quality principles allows improving the analysis, communication and decision processes, as well as sustaining data quality management concepts, methods and tools. Metadata requirements, user experience and perceived quality in spatial data quality evaluation routines based on required field/elements of a specific standard profile implies the enlargement of the ThemisE references, the improvement of the specification and the implementation of new functionalities linked to real context ThemisE evaluations and tests. The work carried out in Task 5.5 allowed advances in SDQ evaluation and management, namely at the level of metadata profiles and the development of the ThemisE platform, providing import contributions to e-science initiatives and linking to global initiatives like GBIF, LIFEWATCH, GEOSS, IPBES and ILTER.

6. References

Agumya, A., and Hunter, G. (2002). Responding to the consequences of uncertainty in geographical data. International Journal of Geographical Information Science, 16(5), 405–417.

Alonso, J., Martins, I., Guerra, C., and Castro, P. (2013). Spatial data quality assessment and WEBGIS institutional capacity building development. In A. Rocha, L., Reis, M., Cota, M. Painho, and M., Neto (Eds.), Sistemas e Tecnologias de Informação- Atas da 8a Conferência Ibérica de Sistemas e Tecnologias de Informação, Vol.I- Artigos. Lisboa, 19 a 22 de Junho: AISTI (Associação Ibérica de Sistemas e Tecnologias de Informação).

Aronoff, S. (1989). Geographical Information Systems: A Management Perspective. Geocarto International, 4(4), 58.

Baranski, B., Foerster, T., Schäffer, B., & Lange, K. (2011). Matching INSPIRE Quality of Service Requirements with Hybrid Clouds. Transactions in GIS, 15(s1), 125–142. https://doi.org/10.1111/j.1467-9671.2011.01265.x

Beare, M. (2010). Data Quality Evaluation - The ESDIN Experience. In Application of Standards for GI ISO/TC211 Interoperability Workshop.

Beare, M., Onstein, E., Mäkelä, J., Marttinen, J., Henriksson, R., Jakobsson, A., Kavadas, I. (2010). D8.4 ESDIN Quality Final Report – Part A- version 1.1.

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

Bobillo, F., Straccia, U., Boin, A. T., Chapman, A. D., Facility, G. B. I., Cotton, D. P., Gahegan, M. (2015). Spatial data quality. International Journal of Geographical Information Science, 85(2), 17. https://doi.org/10.1016/j.jag.2014.10.014

Brassel, K., Bucher, F., Stephan, E., and Vckovski, A. (1995). Completeness. In J. L. Guptill, S.C and Morrison (Ed.), Elements of spatial data quality (pp. 81–108). International Cartographic Association, Tokyo: Elsevier Science.

Castro, P., Alonso, J., Guerra, C., Gonçalves, J., Pôças, I., Marcos, B., and Honrado, J. (2013). Novel Tools to Improve the Management of Spatial Data Quality in the Context of Ecosystem and Biodiversity Monitoring. In Jekel, Adrijana, Strobl, & Griesebner (Eds.), GI_Forum 2013 - Creating the GISociety (Jekel, Adr, pp. 500–503). https://doi.org/doi:10.1553/giscience2013s500

Cuca, B., Brumana, R., Scaioni, M., and Oreni, D. (2011). Spatial Data Management of Temporal Map Series for Cultural and Environmental Heritage. International Journal of Spatial Data Infrastructures Research, 6, 97–125. https://doi.org/10.2902/1725-0463.2011.06.art5

Curdt, C., M, R. M., Ruitton-Allinieu, A.-M., Sidda, N., Ismail, M., Assurance, Q., Stark, H.-J. (2014). Crowdsourcing of Geoinformation: data quality and possible applications. Journal of Industrial Ecology, 7(3–4), 225–226. https://doi.org/10.1162/108819803323059532

Devillers, R., Bédard, Y., and Jeansoulin, R. (2005). Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS. Photogrammetric Engineering & Remote Sensing, 71(2), 205–215.

Devillers, R., Bédard, Y., Jeansoulin, R., and 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.

Devillers, R., & Jeansoulin, R. (2006). Spatial data quality: concepts. In R. Devillers, R., Jeansoulin (Ed.), Fundamentals of Spatial Data Quality (pp. 31–42). ISTE – GIS Series.

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., and Shi, W. (2010). Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. Transactions in GIS, 14(4), 387–400. https://doi.org/10.1111/j.1467-9671.2010.01212.x

Díaz, P., Masó, J., Sevillano, E., Ninyerola, M., Zabala, A., Serral, I., & Pons, X. (2012). Analysis of quality metadata in the GEOSS Clearinghouse. International Journal of Spatial Data Infrastructures Research, 7, 352–377. https://doi.org/10.2902/1725-0463.2012.07.art17

EC. (2010). Data quality in INSPIRE: from requirements to metadata - Data quality and metadata. Contributor: Tóth, K; Tomas, R; Jakobsson, A.; Troispoux, G.; Agius, C.; Nunes de Lima, V.; Smits, P. European Commission.

EC. (2008). Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata. Official Journal of the European Union, 51(L326), 12–30. https://doi.org/citeulike-article-id:11622812

EC. (2010a). COMMISSION REGULATION (EU) No 1089/2010 of 23 November 2010, implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. Official Journal of the European Union L 323/11.

EC. (2010b). Infrastructure for Spatial Information in Europe INSPIRE Metadata Implementing Rules: Changes from V1.1 to V1.2 of Technical Guidelines based on EN ISO 19115 and EN ISO 19119, 1–10.

EC. (2017). Technical Guidelines for implementing dataset and service metadata based on ISO/TS 19139:2007, 178. Retrieved from http://inspire.ec.europa.eu/id/document/tg/metadata-iso19139/2.0.1

Evangelidis, K., Ntouros, K., Makridis, S., and Papatheodorou, C. (2014). Geospatial services in the Cloud. Computers and Geosciences, 63, 116–122. https://doi.org/10.1016/j.cageo.2013.10.007

Gervais, M., Bédard, Y., Levesque, M., Bernier, E., and Devillers, R. (2009). Data Quality Issues and Geographic Knowledge Discovery. In J. Miller, Harvey .J and Han (Ed.), Geographic Data Mining and Knowledge Discovery (Second Edi, pp. 99–115). CRC Press.

Goodchild, M. F. (2007). Beyond Metadata: Towards User- Centric Description of Data Quality. In 5th Int. Symposium Spatial Data Quality. Enschede, Netherlands.

Honrado, J., Alonso, J., Castro, P., Martins, L., Pôças, I., Gonçalves, G., Marcos, B. (2011a). 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. In J., Honrado (Ed.), Project BIO_SOS Biodiversity Multisource Monitoring System: from Space TO Species (p. 123).

Honrado, J., Alonso, J., and Guerra, C. (2011c). WP4 | On-site data collection State of work and Task 4.1. Abery, Wales.

Honrado, J., Alonso, J., Guerra, C., Pôças, I., Gonçalves, J., Marcos, B., and Lucas, R. (2011b). Deliverable No: D4.1- Report on pre-existing in situ and ancillary datasets for sites. In J., Honrado (Ed.), Project BIO_SOS Biodiversity Multisource Monitoring System: from Space TO Species (p. 129).

ISO. (1989). ISO 3951 - Sampling procedures and charts for inspection by variables for percent nonconforming.

ISO. (2002). ISO 19113 - Geographic information - Quality principles.

ISO. (2003). ISO 19114 - Geographic information - Quality evaluation procedures: Draft.

ISO. (2006). ISO 19138 - Geographic information - Data quality measures.

ISO. (2007). ISO 19139 - Geographic information - Metadata - XML schema implementation.

ISO. (2012). ISO 19158 - Geographic information - Quality assurance of data supply. International Organization for Standardization (ISO). Geneva.

ISO. (2013). ISO 19157 - Geographic information - Data quality. International Organization for Standardization (ISO). Geneva.

Ivánová, I., Morales, J., de By, R. A., Beshe, T. S., and Gebresilassie, M. A. (2013). Searching for spatial data resources by fitness for use. Journal of Spatial Science, 58(1), 15–28. https://doi.org/10.1080/14498596.2012.759087

Jakobsson, A. (2011). The European Location Framework – Deployment of EIDEN Results Strategic and Organizational Issues. In EIDEN Workshop. Brussels, Belgium: European Spatial Data Infrastructure Network.

Jakobsson, A., and Giversen, J. (2008). Guideline for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies. Eurogeographics Expert Group on Quality.

Jakobsson, A., and Tsoulos, L. (2007). The Role of Quality in Spatial Data Infrastructures. Eurogeographics, 10. ICC Proceedings, ICC2007

Leibovici, D. G., Pourabdollah, A., and Jackson, M. J. (2013). Which spatial data quality can be meta-propagated? Journal of Spatial Science, 58(1), 3–14. https://doi.org/10.1080/14498596.2012.759088

Lopez-Pellicer, F., Béjar, R., Florczyk, A., Muro-Medrano, P., and Zarazaga-soria, F. (2011). A Review of the Implementation of OGC Web Services across Europe. International Journal of Spatial Data Infrastructures Research, 6, 168–186. https://doi.org/10.2902/1725-0463.2011.06.art8

Lush, V., Nüst, D., Bastin, L., Masó, J., and Lumsden, J. (2014). GEO Label Web Services for Dynamic and Effective Communication of Geospatial Metadata Quality, 16, 14453.

Manso-Callejo, M., Wachowicz, M., and Bernabé-Poveda, M. (2008). Automatic Metadata Creation for Supporting Interoperability Levels of Spatial Data Infrastructures.

National Land & Water Resources Audit. (2008). Natural Resources Information Management Toolkit: building capacity to implement natural resources information management solutions. Canberra.

Ninyerola, M., Sevillano, E., Serral, I., Pons, X., Zabala, A., and Bastin, L. (2014). QualityML: a dictionary for quality metadata encoding, 16, 19157.

OGC. (2007). OpenGIS ® Catalogue Services Specification. (P. Nebert, D; Whiteside, Arliss; Vretanos, Ed.) (Version 2.). Open Geospatial Consortium Inc.

Olfat, H. (2013). Automatic Spatial Metadata Updating and Enrichment. PhD Thesis. The University of Melbourne, Victoria, Australia.

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H., and Williamson, I. (2012). Spatial Metadata Automation: A Key to Spatially Enabling Platform. International Journal of Spatial Data Infrastructures Research, 7, 173–193. https://doi.org/10.2902/1725-0463.2012.07.art9

Pôças, I., Gonçalves, J., Marcos, B., Alonso, J., Castro, P., and Honrado, J. P. (2014). Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring. International Journal of Geographical Information Science, 28(11), 2356–2371. https://doi.org/10.1080/13658816.2014.924627

Pons, X., and Masó, J. (2016). A comprehensive open package format for preservation and distribution of geospatial data and metadata. Computers and Geosciences, 97(September 2015), 89–97. https://doi.org/10.1016/j.cageo.2016.09.001

Sebake, M., and Coetzee, S. (2013). Address Data Sharing: Organizational Motivators and Barriers and their Implications for the South African Spatial Data Infrastructure. International Journal of Spatial Data Infrastructures Research, 8. https://doi.org/10.2902/1725-0463.2013.xx.arty

Tóth, K., Tomas, R., Jakobsson, A., Troispoux, G., and Agius, C. (2010). Infrastructure for Spatial Information in Europe Data quality in INSPIRE: from requirements to metadata, 1–21.

van Oort, P. (2006). Spatial data quality: from description to application. Delft: Publications on Geodesy 60, NCG, Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission.

Van Oort, P. A. J., and Bregt, A. K. (2005). Do users ignore spatial data quality? A decision-theoretic perspective. Risk Analysis, 25(6), 1599–1610. https://doi.org/10.1111/j.1539-6924.2005.00678.x

Vasseur, B., Devillers, R., and Jeansoulin, R. (2003). Ontological Approach of the Fitness of use of Geospatial Datasets. In Proceedings of the 6th AGILE (pp. 497–504). Lyon-France.

Victorian Spatial Council. (2009). Spatial Information Business Information Guidelines. In Part of Victoria's Spatial Information Management Framework (2nd Edition). East Melbourne: Victorian Spatial Council C/-Spatial Information Infrastructure.

Waters, R., Beare, M., Walker, R., and Millot, M. (2011). Schema Transformation for INSPIRE. International Journal of Spatial Data Infrastructures Research, 6, 1–22. https://doi.org/10.2902/1725-0463.2011.06.art1

7. Appendices

Annex I – Questionnaire model on ”Knowledge and Routines of Data Quality Assessment and Management”

Annex II – Results of questionnaire ”Knowledge and Routines of Data Quality Assessment and Management”

Annex III – THEmatic Metadata-based and fItness for use Spatial data quality Evaluation: User manual

7.1 Annex I – Questionnaire model on "Knowledge and Routines of Data Quality Assessment and Management"

- **Online questionnaire** "KNOWLEDGE AND ROUTINES OF DATA QUALITY ASSESSMENT AND MANAGEMENT"

Introduction

ECOPOTENTIAL: improving future ecosystem benefits through earth observations
WP5 – In situ Monitoring Data (WP Leader: EAA)
Task 5.3 – Provision of consistent and standard compliant metadata (Lead: FORTH)
Task 5.5 – Development of data quality evaluation routines (Lead: ICETA/CIBIO)

The questionnaire on "knowledge and routines of data quality assessment and management" aims to inquire the ECOPOTENTIAL community of data providers and data users about their KNOWLEDGE, their PRACTICAL EXPERIENCE and their AWARENESS (UTILITY) of data quality evaluation routines/tools. It consists of a collaborative and oriented online questionnaire with closed and open questions.

The questionnaire is aimed to inquire individual Ecopotential researchers, including data providers and data users, on quality issues related to EO data production and usage, field data production and usage, and application of data ecological models. More specifically: (1) knowledge of data quality, including quality theory, concepts, elements (KNOWLEDGE); (2) current practices (e.g. practical experience) related to (spatial) data quality assessment and management routines (PRACTICAL EXPERIENCE); and (3) awareness, interests and willingness to implement data quality routines (UTILITY).

The questionnaire results will be of high relevance for future activities in WP5, namely when: (1) defining the fields/attributes of quality elements to include in the selected metadata profile (Task 5.3); (2) specifying and developing methods (external evaluation) and tools (routines and information/ technological application) of spatial data quality assessment (Task 5.5); and (3) devising proposals for implementation of quality management processes (WP5).

Answering the questionnaire will take you approx. 10/15 minutes.
Responses should be submitted until **31 March 2017** (*second reminder **14th April 2017;** third reminder **31st May 2017***)

- **Issues** (structure of online questionnaire)

(1) IDENTIFY THE SPATIAL DATA USER'S CHARACTERISTICS AND THEIR POSITION WITH SPATIAL DATA LIFE CYCLE?

Academic background
( ) Biology/Ecology
( ) Mathematics/Statistics
( ) Engineering/Technology
( ) Geophysics/Geography
( ) Human and Social Sciences
( ) Remote Sensing/Spatial data

( ) Others

Researcher with
( ) Degree in ……………………….. University(ies) ……………………………..
( ) Master …………………………… University(ies) ……………………………..
( ) PhD ……………………………… University(ies) ……………………………..
( ) Post PhD ……………………………… University(ies) ……………………………..
Describe your academic degrees and institution (e.g. Degree in Biology - UP)

Researcher in Partner
Italy: CNR, UNILE, EURAC, ISPRA, POLIMI
Spain: CSIC, CREAF, UAB, IISTA-UGR, STARLAB, REDIAM
Germany: UFZ, KIT, UBT, DLR, UPotsdam, MfN, iDiv-MLU
France: CNRS-UMR, TdV, CESBIO-UPS, UBO
UK: UNIVLEEDS, ESL, LSE, UKT2
Romania: UB
Portugal: ICETA, IST
Greece: CERTH, FORTH, ARATOS
Switzerland: EPFL, ETH, UNIGE
Israel: BGU, INPA
Macedonia: HIO
South Africa: CSIR
Austria: EAA
Netherlands: DELTARES, NIOZ
Lithuania: KU
Norway: UiB
Intern. Entity: UNESCO, UNEP
Australia: UNSW
Hungray: *SIU*
Sweden: *UUmea*
Venezuela: *Provita/IUCN*

Researcher in WP (one or more selection)
( ) WP1 - Coordination and management
( ) WP2 - Conceptual Scientific Framework
( ) WP3 - Earth Observation Data and Processes Infrastructure
( ) WP4 - Earth Observation Data Generation and Harmonization
( ) WP5 - In situ Monitoring Data

( ) WP6 - EO-based Ecosystem Modelling
( ) WP7 - Ecosystem Services
( ) WP8 - Cross-scale interaction
( ) WP9 - Requirements of future protected areas
( ) WP10 - ECOPOTENTIAL Virtual Laboratory Platform
( ) WP11 - EO supported policy development and integration
( ) WP12 - Capacity building and knowledge exchange

Classify your position relating spatial data life cycle
( ) Data provider
( ) Data user
( ) Mainly data provider
( ) Mainly data user
( ) Both (equilibrium data provider/supplier and data user/consumer)

(2) SPATIAL (META)DATA KNOWLEDGE

(2.1) User knowledge about (spatial) data quality/ Domain of data quality concepts (for geographic information or spatial data)

Knowledge about ISO standards associated with the spatial data quality?

	No	Yes	Fair	Very Low	Low	Medium	Good	VeryGood	Excellent
ISO19113:2002 - Quality principles, and establishes the principles for describing the quality of geographic data. https://www.iso.org/standard/26018.html	o	o	o	o	o	o	o	o	o
ISO19114:2003 - Quality evaluation procedures. https://www.iso.org/standard/26019.html	o	o	o	o	o	o	o	o	o
ISO19138:2006 - Data quality measures. https://www.iso.org/standard/32556.html	o	o	o	o	o	o	o	o	o
ISO19157:2013 - Data quality (DQ) for geographic data. https://www.iso.org/standard/32575.html	o	o	o	o	o	o	o	o	o
ISO19158:2012 - Quality assurance of data supply. https://www.iso.org/standard/32576.html	o	o	o	o	o	o	o	o	o

Additional notes/Comments
Other: ____

(2.2) User knowledge about spatial data quality assessment and management

Do you know the following metadata standards/application schemas?

	No	Yes	Fair	VerLow	Low	Medium	Good	Very Good	Excellent
INSPIRE MS Specification http://inspire.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf	o	o	o	o	o	o	o	o	o
ISO19115-1:2014 - Metadata Part1: Fundamentals. https://www.iso.org/standard/53798.html	o	o	o	o	o	o	o	o	o
ISO 19115-2:2009 - Metadata Part 2: Extensions for imagery and gridded data. https://www.iso.org/standard/39229.html									
ISO19139:2007 - Metadata XML schema implementation. https://www.iso.org/obp/ui/#iso:std:iso:ts:19139:ed-1:v1:en	o	o	o	o	o	o	o	o	o
DEIMS (Dynamic Ecological Information Management System) https://data.lter-europe.net/deims/documentation/dataset	o	o	o	o	o	o	o	o	o
EML (Ecological MD Language) http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language	o	o	o	o	o	o	o	o	o
DC (Dublin Core) http://dublincore.org/	o	o	o	o	o	o	o	o	o
DwC (Darwin Core) http://www.tdwg.org/activities/darwincore/	o	o	o	o	o	o	o	o	o
ISO19157:2013 - Data quality (DQ) for geographic data. https://www.iso.org/standard/32575.html	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Other: ____

(2.3) User knowledge about data quality elements?

Do you know data quality elements? Data Quality elements (ISO19157:2013)

	No	Yes	Fair	VeryLow	Low	Medium	Good	Very Good	Excellent
Completeness (presence and absence of features, their attributes and their relationships)	o	o	o	o	o	o	o	o	o
Logical Consistency (degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical))	o	o	o	o	o	o	o	o	o
Positional Accuracy (accuracy of the position of features)	o	o	o	o	o	o	o	o	o
Thematic Accuracy (accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships)	o	o	o	o	o	o	o	o	o
Temporal Quality (accuracy of the temporal attributes and temporal relationships of features)	o	o	o	o	o	o	o	o	o
Usability Element (degree of adherence of a dataset to a specific set of requirements)	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Other: ____

(2.4) User knowledge about spatial data quality management

Do you know and/or you are implemented procedures of spatial data quality management along spatial data life cycle?

	No	Yes	Fair	VeryLow	Low	Medium	Good	VeryGood	Excellent
Data product specification (e.g. ISO19131:2007 - Data product specifications, geographic data https://www.iso.org/standard/36760.html)	o	o	o	o	o	o	o	o	o
Data product specification user requirements and provide quantitative quality information (e.g. data models on Environmental Monitoring Facilities, Species distribution, Habitats and biotopes... http://inspire.ec.europa.eu/data-specifications/2892)	o	o	o	o	o	o	o	o	o
Collecting data (data collection protocols) (e.g. ISO 2854:1976 - Statistical interpretation of data https://www.iso.org/standard/7854.html; ISO 3534-1:2006 - Statistics Part 1: General statistical terms and terms used in probability https://www.iso.org/standard/40145.html; ISO 3534-4:2014 - Statistics Part 4: Survey sampling https://www.iso.org/standard/56154.html)	o	o	o	o	o	o	o	o	o
Instrumental/technological component (e.g. instrumental selection and calibration)	o	o	o	o	o	o	o	o	o
Individual and organizational capacity building (Human component at specification and implementation methods) (e.g. ISO 19122:2004 - Qualification and certification of personnel https://www.iso.org/standard/31088.html)	o	o	o	o	o	o	o	o	o
Data conceptual modelling (e.g. ISO19109:2015 - Rules for application schema, geographic data https://www.iso.org/standard/59193.html)	o	o	o	o	o	o	o	o	o
Spatial analysis and modelling (e.g. data quality evaluation on Correlative and Process-based models)	o	o	o	o	o	o	o	o	o
Data representation, publishing and sharing (e.g. analogic and digital graphical data communication relating data dissemination/accessibility)	o	o	o	o	o	o	o	o	o
Evaluating data quality/Reporting data quality (spatial data quality evaluation/metadata) (e.g. ISO19157:2013 - Data quality (DQ) for geographic data. https://www.iso.org/standard/32575.html)	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Others: ____

(3) SPATIAL (META)DATA QUALITY INTEREST

(3.1) Do you consider or you are interested in know/use spatial data quality elements?

No	Yes	Fair Interest	Very Low	Low	Medium	High	Very High	Total interest
o	o	o	o	o	o	o	o	o

(3.2) Do you consider or you are interested in know/use spatial data quality assessment methods and tools?

No	Yes	Fair Interest	Very Low	Low	Medium	High	Very High	Total interest
o	o	o	o	o	o	o	o	o

(3.3) Do you consider or you are interested in use/participate in spatial data quality management process?

No	Yes	Fair Interest	Very Low	Low	Medium	High	Very High	Total interest
o	o	o	o	o	o	o	o	o

(4) SPATIAL (META)DATA QUALITY EXPERIENCE

On your activities as data provider or data user have you applied data quality evaluation practices/routines?

No	Yes	Very Rarely	Rarely	Sometimes	Occasionally	Regularly	Often	Always
o	o	o	o	o	o	o	o	o

(4.1) Which **measures** (the type of evaluation) you applied?

	No	Yes	Very Rarely	Rarely	Sometimes	Occasionally	Regularly	Often	Always
Qualitative measures (e.g. descriptive)	o	o	o	o	o	o	o	o	o
Quantitative measures (e.g. number of excess items)	o	o	o	o	o	o	o	o	o
Qualitative/quantitative measures	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Others: ____

(4.2) Which **evaluation methods** (the procedure used to evaluate the measure) has applied?

	No	Yes	Very Rarely	Rarely	Sometimes	Occasionally	Regularly	Often	Always
Direct internal (method of evaluating the quality of a dataset based on inspection of items within the dataset,	o	o	o	o	o	o	o	o	o

where all data required is internal to the dataset being evaluated)									
Direct external (method of evaluating the quality of a dataset based on inspection of items within the dataset, where reference data external to the dataset being evaluated is required)	○	○	○	○	○	○	○	○	○
Indirect (method of evaluating the quality of a dataset based on external knowledge)	○	○	○	○	○	○	○	○	○
Aggregation and derivation methods	○	○	○	○	○	○	○	○	○

Additional notes/Comments

Others: ____

(4.3) Which **results** obtained in your spatial data quality evaluation process?

	No	Yes	Very Rarely	Rarely	Sometimes	Occasionally	Regularly	Often	Always
Quantitative results (quantitative result may be a single value or multiple values, depending on the values of attributes defined in the description of the measure applied)	○	○	○	○	○	○	○	○	○
Conformance results (a conformance result is the outcome of comparing the value or set of values obtained from applying a measure to the data specified by a data quality scope with a specified acceptable conformance quality level)	○	○	○	○	○	○	○	○	○
Descriptive results (in some cases (e.g. with thematic and geoscientific observations), it is not possible to produce a quantitative result for a data quality element. A subjective evaluation of an element can then be expressed with a textual statement as a data quality descriptive result)	○	○	○	○	○	○	○	○	○

Additional notes/Comments

Others: ____

(5) SPATIAL (META)DATA QUALITY UTILITY

(5.1) Do you consider or communicate in **quality assurance** (QA)/ **quality control** (QC):

Do you implement procedures for data **quality assurance** (QA)/ **quality control** (QC)?

No	Yes	Very Rarely	Rarely	Sometimes	Occasionally	Regularly	Often	Always
○	○	○	○	○	○	○	○	○

Do you implement and publish documentation about results of QA/QC procedures in metadata?

No	Yes	Very Rarely	Rarely	Sometimes	Occasionally	Regularly	Often	Always
o	o	o	o	o	o	o	o	o

(5.2) Which **data quality elements** (ISO19157:2013) do you consider decisive:

To discover and select input data for applying models and workflows?

	No	Yes	Fair	Very Low	Low	Medium	High	Very High	Decisive
Completeness	o	o	o	o	o	o	o	o	o
Logical Consistency	o	o	o	o	o	o	o	o	o
Positional Accuracy	o	o	o	o	o	o	o	o	o
Thematic Accuracy	o	o	o	o	o	o	o	o	o
Temporal Quality	o	o	o	o	o	o	o	o	o
Usability Element	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Others: ____

To explore the results of practical/ecological meaning of output data?

	No	Yes	Fair	Very Low	Low	Medium	High	Very High	Decisive
Completeness	o	o	o	o	o	o	o	o	o
Logical Consistency	o	o	o	o	o	o	o	o	o
Positional Accuracy	o	o	o	o	o	o	o	o	o
Thematic Accuracy	o	o	o	o	o	o	o	o	o
Temporal Quality	o	o	o	o	o	o	o	o	o
Usability Element	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Others: ____

To communicate with end user/technical-political decision makers?

	No	Yes	Fair	Very Low	Low	Medium	High	Very High	Decisive
Completeness	o	o	o	o	o	o	o	o	o
Logical Consistency	o	o	o	o	o	o	o	o	o
Positional Accuracy	o	o	o	o	o	o	o	o	o

Thematic Accuracy	o	o	o	o	o	o	o	o	o
Temporal Quality	o	o	o	o	o	o	o	o	o
Usability Element	o	o	o	o	o	o	o	o	o

Additional notes/Comments

Others: ____

(5.3) Which data quality elements/indicators do you considers important/relevant to incorporate into a metadata profile?

Elements/indicators:

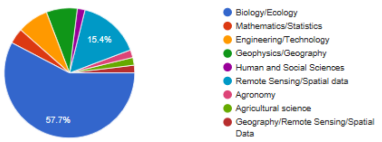
( ) **Typology** (topic category defining the main data set theme)
( ) **Taxonomic coverage** (taxonomic classification of the organisms represented in the dataset)
( ) **Completeness Commission** (number of excess items)
( ) **Completeness Omission** (number of missing items)
( ) **Conceptual consistency** (number of items not compliant with the rules of the conceptual schema)
( ) **Spatial extent** (bounding box defining spatial coverage)
( ) **Temporal extent** (time interval defining temporal coverage)
( ) **Lineage** description in metadata (description of data production methods and overall quality)
( ) **Methods description** (provides repeated sets of elements that document a series of procedures followed to produce any dataset object)
( ) **Instrumentation description** (provides information about any instruments used in the data collection or quality control and quality assurance)
( ) **Sampling description** (provides information about sampling part of the method as measurement frequency, and spatial scale)
( ) **Quality assurance** (provides information on QA/QC procedures applied for the data)
( ) **Legal obligation reporting** (provides information whether the dataset has been reported to the local, regional or national bodies to fulfil the obligations from particular legal regulations)
( ) **Thematic accuracy** (data set thematic accuracy; e.g. number of incorrectly classified features; kappa coefficient)
( ) **Spatial scale** (equivalent scale or spatial resolution defining the level of detail)
( ) **Temporal quality** (accuracy of the temporal attributes and temporal relationships of features)
( ) **Producer recognition** (data producer recognition type)
( ) **Intellectual Rights** (list of rights management statements for the dataset, or reference a URL (web address) that provides such information)
( ) **Access and use constraints** (conditions applying to access and use)
( ) **File format** (distribution file format)
( ) **Online distribution** (web address is the "navigation section" of a metadata record pointing users to the location (URL) where a dataset can be retrieved directly, or provides information about how to acquire a dataset)
( ) **Usability** (degree of adherence of a dataset to a specific set of requirements)
Others: ____

7.2 Annex II – Results of questionnaire ”Knowledge and Routines of Data Quality Assessment and Management”

(1) Identify the spatial data user’s characteristics and their position with spatial data life cycle?

Academic background

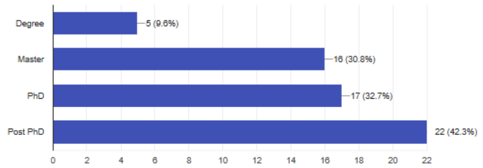
52 responses



(1) Identify the spatial data user’s characteristics and their position with spatial data life cycle?

Researcher with

52 responses



University(ies)

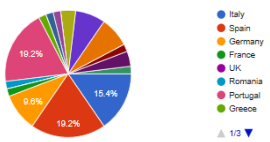
52 responses

- Ecology and environmental science
- All in Ecology at Klaipeda university
- BSc. and MSc. in Geography with focus on remote sensing and land-use mapping
- PhD in Biology
- PhD in Ecology
- PhD in Biology - Mediterranean Institute for Advanced Studies (Spanish Research Council CSIC)
- degree in cartography - University
- PhD in Biology - University of Porto, Portugal
- Degree and MSc in Environmental Science - UP
- Degree in Sciences of Engeneering - UP
- Degree in Biology - UAM, PostGraduate in Advanced Environmental Ecotechnology - WAU & IHE-Delft, PhD in Agricultural & Environmental Sciences - WAU

(1) Identify the spatial data user’s characteristics and their position with spatial data life cycle?

Researcher in Partner

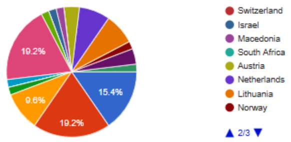
52 responses



(1) Identify the spatial data user’s characteristics and their position with spatial data life cycle?

Researcher in Partner

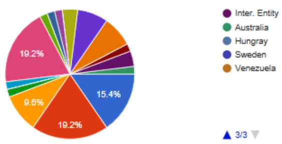
52 responses



(1) Identify the spatial data user’s characteristics and their position with spatial data life cycle?

Researcher in Partner

52 responses



Researcher in Partner - Lithuania

4 responses

Researcher in Partner - Norway

1 response

Researcher in Partner - Switzerland

0 responses

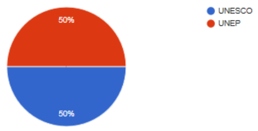
No responses yet for this question.

Researcher in Partner - Macedonia

1 response

Researcher in Partner - Intern. Entity

2 responses



Researcher in Partner - South Africa

0 responses

No responses yet for this question.

Researcher in Partner - Australia

1 response

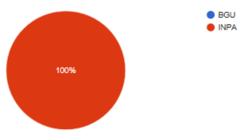
Researcher in Partner - Venezuela

0 responses

No responses yet for this question.

Researcher in Partner - Israel

1 response



Researcher in Partner - Sweden

0 responses

No responses yet for this question.

Researcher in Partner - Hungray

0 responses

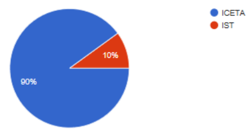
No responses yet for this question.

Researcher in Partner - Romania

1 response

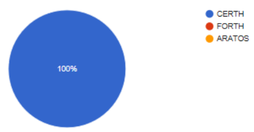
Researcher in Partner - Portugal

10 responses



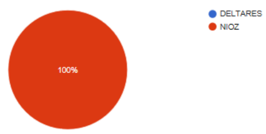
Researcher in Partner - Greece

1 response



Researcher in Partner - Netherlands

4 responses



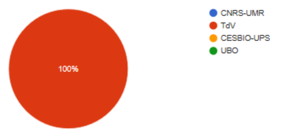
Researcher in Partner - UK

0 responses

No responses yet for this question.

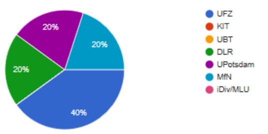
Researcher in Partner - France

1 response



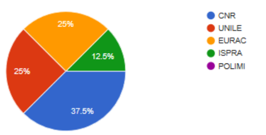
Researcher in Partner - Germany

5 responses



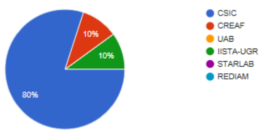
Researcher in Partner - Italy

8 responses



Researcher in Partner - Spain

10 responses



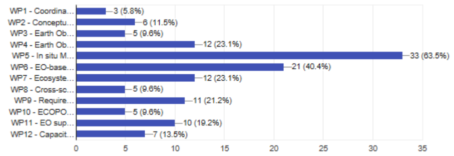
Researcher in Partner - Austria

2 responses

(1) Identify the spatial data user's characteristics and their position with spatial data life cycle?

Researcher in WP (one or more selection)

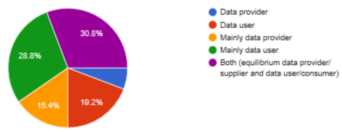
52 responses



(1) Identify the spatial data user's characteristics and their position with spatial data life cycle?

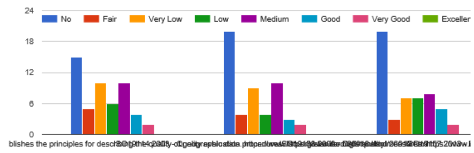
Classify your position relating spatial data life cycle

52 responses



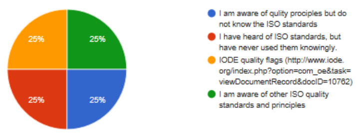
(2) Spatial (meta)data knowledge

Knowledge about ISO standards associated with the spatial data quality?



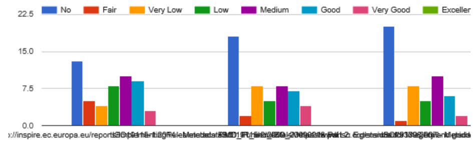
Additional notes/Comments

4 responses



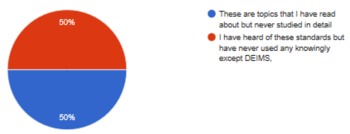
(2) Spatial (meta)data knowledge

Do you know the following metadata standards/application schemas?



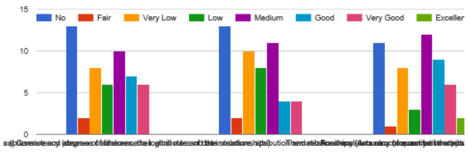
Additional notes/Comments

2 responses



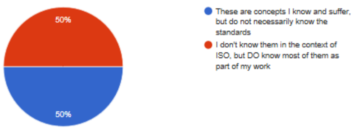
(2) Spatial (meta)data knowledge

Do you know data quality elements? Data Quality elements (ISO19157:2013)



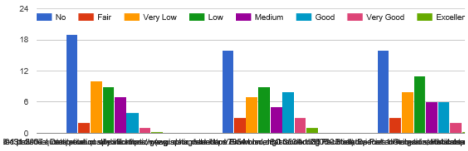
Additional notes/Comments

2 responses



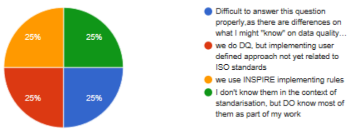
(2) Spatial (meta)data knowledge

Do you know and/or you are implemented procedures of spatial data quality management along spatial data life cycle?



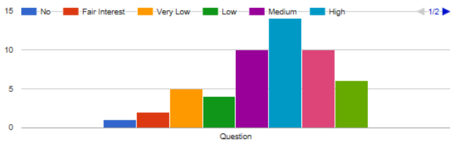
Additional notes/Comments

4 responses



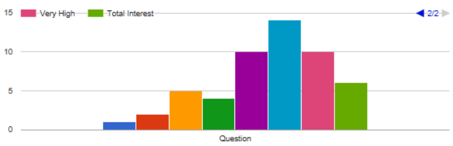
(3) Spatial (meta)data quality interest

(3.1) Do you consider or you are interested in know/use spatial data quality elements?



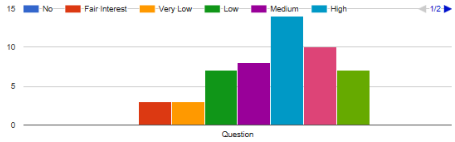
(3) Spatial (meta)data quality interest

(3.1) Do you consider or you are interested in know/use spatial data quality elements?



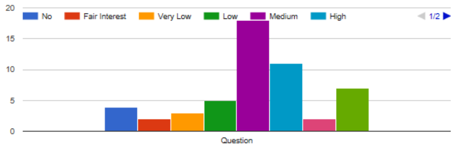
(3) Spatial (meta)data quality interest

(3.2) Do you consider or you are interested in know/use spatial data quality assessment methods and tools?



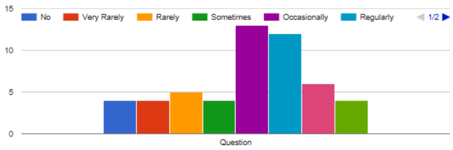
(3) Spatial (meta)data quality interest

(3.3) Do you consider or you are interested in use/participate in spatial data quality management process?



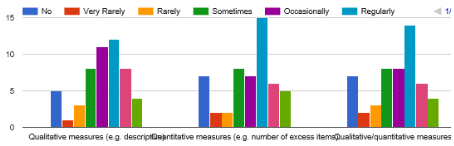
(4) Spatial (meta)data quality experience

On your activities as data provider or data user have you applied data quality evaluation practices/routines?



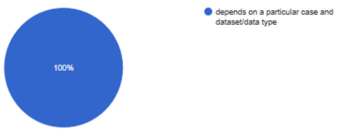
(4) Spatial (meta)data quality experience

(4.1) Which measures (the type of evaluation) you applied?



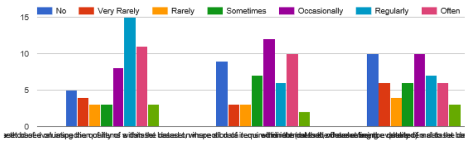
Additional notes/Comments

1 response



(4) Spatial (meta)data quality experience

(4.2) Which evaluation methods (the procedure used to evaluate the measure) has applied?

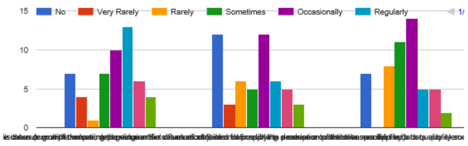


Additional notes/Comments

1 response

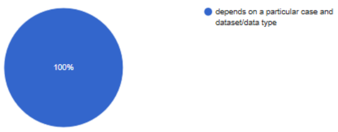
(4) Spatial (meta)data quality experience

(4.3) Which results obtained in your spatial data quality evaluation process?



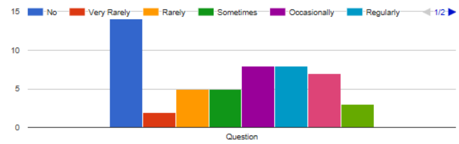
Additional notes/Comments

1 response



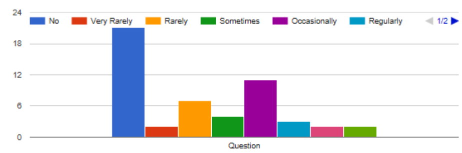
(5) Spatial (meta)data quality utility

Do you implement procedures for data quality assurance (QA)/ quality control (QC)?



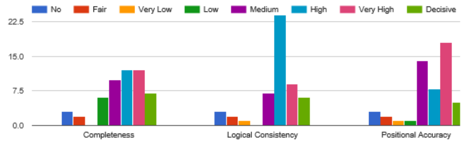
(5) Spatial (meta)data quality utility

Do you implement and publish documentation about results of QA/QC procedures in metadata?



(5) Spatial (meta)data quality utility

To discover and select input data for applying models and workflows?

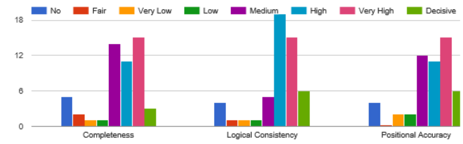


Additional notes/Comments

1 response

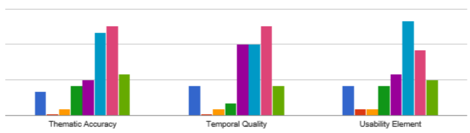
(5) Spatial (meta)data quality utility

To explore the results of practical/ecological meaning of output data?



(5) Spatial (meta)data quality utility

To explore the results of practical/ecological meaning of output data?

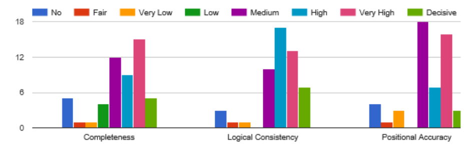


Additional notes/Comments

1 response

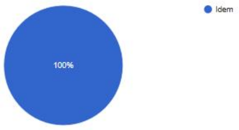
(5) Spatial (meta)data quality utility

To communicate with end user/technical-political decision makers?



Additional notes/Comments

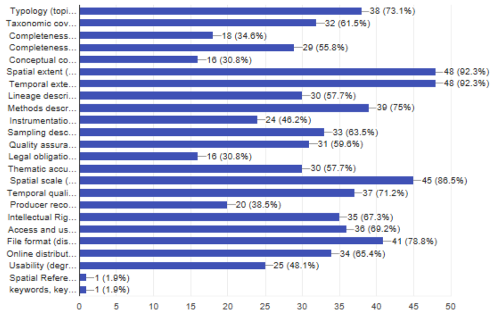
1 response



(5) Spatial (meta)data quality utility

Elements/Indicators

52 responses



7.3 Annex III – THEmatic Metadata-based and fItness for use Spatial data quality Evaluation: User manual

ThemisE platform:

User Manual

Table of Contents

Introduction

The ThemisE platform has been developed in the context of the "ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS" project. This platform has been implemented as a Web application, accessible through any modern browser. ThemisE aims to support the evaluation and identification of relevant data for specific user´s application contexts based on the use of user-oriented quality evaluation routines and procedures for the assessment of the quality of pre-existing data based on (spatial) metadata. The ThemisE platform provides a user-friendly interface with the needed functionalities to perform the evaluation of datasets in terms of their adequacy according to user's data requirements (i.e., the characteristics of the data as defined by the user). The evaluation process requires the user to: (i) specify the targeted themes and their quality indicators (i.e., the characteristics of the themes to search for), (ii) specify/select the metadata database catalogues to be used as the source of datasets to be evaluated, and, (iii) execute the evaluation routines to be able to visualize the results.

In this context, the following sections of the user's manual describes the functionalities of the user interface of the ThemisE platform related to: (i) the main user graphical interface; (ii) the necessary actions to define the user's requirements to be used in the (meta)data quality evaluation process; (iii) the execution of the evaluation; (iv) the visualization of evaluation results and (v) the options to save/load the data introduced by the user.

Main Interface

The main interface of the ThemisE platform consists of a Web page, accessible using a browser, with a grid layout and a toolbar at the bottom left side (Figure 1), that provides the necessary tools to set the user's requirements used to evaluate the datasets for a specific application context and to define and select the sources of metadata database catalogues.

Thus, the evaluation setup consists of two main steps. The first one involves the definition of expected values (Figure 1 – item 1) for each relevant quality indicator (Figure 1 – item 2), where multiple targeted datasets (organized by Thematic Categories) can be configured (Figure 1 – item 3). The second step of the configuration requires the specification of the Datasets sources (Figure 1 – item 4) through the configuration of metadata catalogues that can be used the evaluation. Once this configuration is finished, the evaluation routines can be started. In the next subsections, the necessary actions to complete each one of these steps are described in more detail.

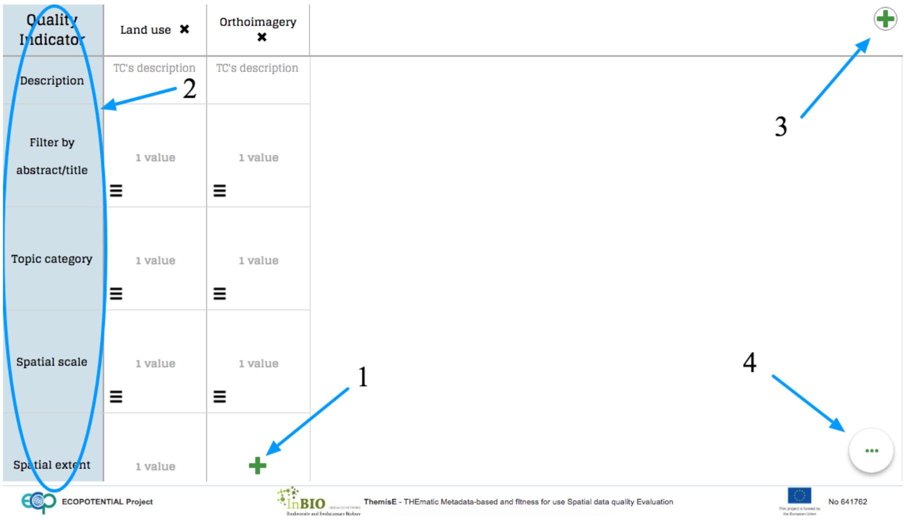


Figure 1 - Main page of the ThemisE platform.

Evaluation

Setup of targeted datasets/Quality indicators

The setup of targeted datasets is made through the definition of the expected values for relevant quality indicators (for each targeted dataset), considering the specificity of each dataset type and the application context. In general, the specification of each quality indicator includes the logical combination of expected values and the option to set it as a critical factor and/or as a filter through the grid layout.

The grid page is organized in columns that correspond to datasets to search/identify (targeted datasets) and lines corresponding to the available and configured quality indicators (Figure 1). Each grid cell offers the tools needed for the definition of the expected values for a particular quality indicator (line) of a specific targeted dataset (column). Additionally, each cell displays a summary text with information of the number of specified values.

To begin the setup of targeted datasets the user must insert the targeted datasets he/she wants to evaluate. The insertion of each targeted dataset is made through the use the “plus” icon on the top-right (Figure 1 – item 3). A popup window will show up with an input box where the user can select the thematic category (TC) of the targeted dataset and click on the “Add” button. This operation will result in the adding of a new column with a title corresponding to the selected TC. In order to facilitate the organization of the different targeted datasets, the user has the possibility to insert a description (as free text) by clicking on the first grid line of the inserted column (line identified as *‘Description’*). Note that this step must be repeated in order to add as many columns as targeted datasets.

After the specification of the targeted datasets, the user has to specify the expected data quality values for the applicable quality indicators of each targeted dataset corresponding to the cells of the grid. This process must be repeated as many times as necessary to define all the datasets quality characteristics that are needed for the case in analysis. The initial configuration of each pair quality indicator/targeted dataset is made through the use of the “plus” button on the pretended cell (Figure 1 – item 1). This action will result in the opening of a form for the entry/editing of expected quality values according to the quality indicators’ data types and respective rules to combine multiple values, allowing the definition of more advanced conditions. These forms present several aspects common to all quality indicators. The Figure 2 presents an example of the layout of a quality indicator configuration, with two main sections: (i) the definition of expected values with comparisons rules, and, (ii) options.

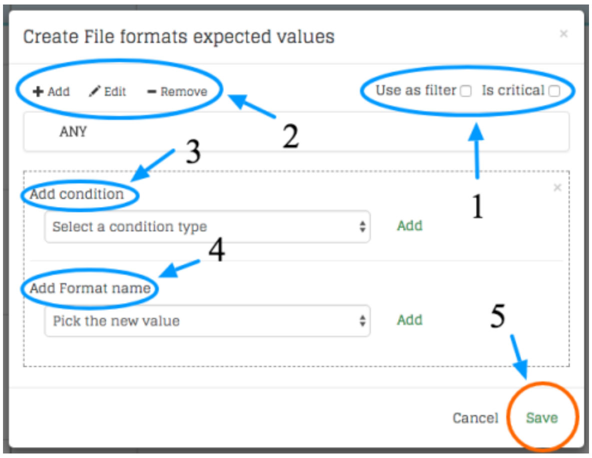


Figure 2 – Form to specify expected values for a quality indicator.

The definition of expected values with comparisons rules of each quality indicator is made through three main actions: ‘Add’, ‘Edit’ and ‘Remove’ (Figure 2 – item 2). To be able to use the ‘Add’ action, the user must first select a row of the table of values and then click on “Add” button to define a new pair of Condition (Figure 2 – item 3)/Value (Figure 2 – item 4) that will be presented as a new line in the table below the ‘Add’ action. The specification of expected values (Figure 2 – item 4) will be adapted to each quality indicator type, and can be made through the entry of free text or numbers, the choice of item(s) from (multiple) lists, the selection of a time interval, the definition of a bounding box for geographic area, etc. Regarding the condition, two are available: *ALL* to specify that all values must be verified and, *ANY* to indicate that at least one defined value must be matched when comparing expected values with metadata values. The ‘Edit’ and ‘Remove’ actions allows editing or removing a selected line of the table of expected values.

Through the adding/editing/removing of multiple values combined through different conditions allows defining multiple expected values that are presented as a multi-level table where each level corresponds to a rule (condition) with respective values. For example, Figure 3 presents a multi-value definition for the topic category, specifying that a quality indicator is conformant if the metadata element has classification with value (“Inland Waters” or “Oceans”) and value (“Transportation” or “Utilities/Communications”).

Figure 3 – Example of the definition of multi-values of expected values.

Besides the specification of expected values, the user has the possibility to select some options (Figure 2 – item 1), depending on the quality indicators. The available options are presented on the top-right corner of the form (Figure 2 – item 1):

- 'Use as filter': when this option is checked, all the datasets that do not verify the defined expected values for the quality indicator will be excluded from the final results. Therefore, this option allows to filter (eliminate) datasets that are not relevant in order to facilitate the analysis process;
- 'Is critical': used to specify the if quality indicator conformity is critical. Unlike the previous option, the datasets with this option activated that do not conform will be presented in the final results as non-conformant. The use of this option over the previous one may allow the user to identify requirements that are crucial to attain;
- 'Cover (%)': this option is available for quality indicators with an extent definition such as for spatial bounding-boxes and temporal extents. This option has a default value of 100% that can be lowered to specify the percentage of coverage that is required to classify the quality indicator as conformant.

After completing the expected values in the form, the user must click on the 'Save' button (Figure 2 – item 5) to store the introduced configuration.

After completing the definition of expected values for a pair quality indicator/targeted dataset, the corresponding cell of the grid will display the number of inserted expected values allowing to visualize the completion degree of the grid (Figure 4 presents the display of a cell before and after the configuration is done).

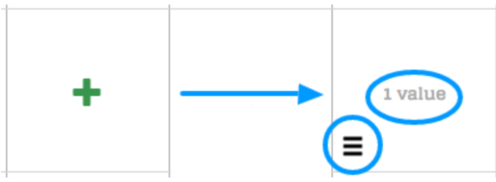


Figure 4 - Cell before (left) and after (right) quality indicator specification.

In addition, a new menu is made available within the cell with three functionalities: *'Remove configuration'*, *'Edit configuration'* and *'Copy configuration from'* (from left to right in Figure 5):

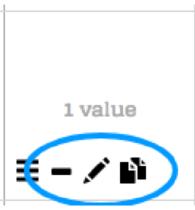


Figure 5 - Cell's functionalities.

- 'Remove *configuration'*: allows the user to remove the cell's configuration (all defined expected values will be removed);
- *'Edit configuration'*: opens a form allowing the user to edit the expected values and conditions defined for the selected pair quality indicator/targeted dataset;
- *'Copy configuration from'*: allows the user to copy expected values from another column (targeted dataset) of the same row (quality indicator). For example, this option can be used to copy expected values of spatial extent quality indicators between different targeted datasets, avoiding the repeated introduction of the same values for each targeted dataset.

The setup of each targeted dataset has a 'special' row identified as *'Filter by abstract/title'* in the grid (Figure 6), as it is not a quality indicator (the setting of this cell will not be used in the evaluation). This option has been added to cope with catalogues with a large number of datasets thus allowing the user to define values to be verified in the abstract/title element of the metadata, and to be used as a filter to exclude from the evaluation process the datasets that are not valid.

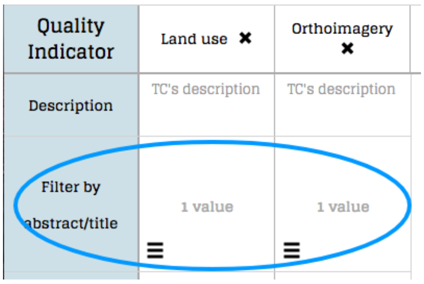


Figure 6 – Option to 'Filter by abstract/title'.

Configure datasets catalogues

In order to be able to perform the evaluation, the user has to setup, at least, one metadata database catalogue source. From this, the platform will get metadata records to be evaluated against the specified quality indicators (previous section).

The setting of a metadata database catalogue is made through the selection of the option 'Settings' on the bottom-right menu (Figure 1 – item 4) of the main page, which gives access to the configuration window of dataset sources (figure 7). This window includes functionalities to add, duplicate and remove catalogue sources (Figure 7 – item 1).

The evaluation, to be successfully started, needs to have at least one datasets catalogue, to accomplish this, we must access the option *'Settings'* on the bottom-right menu (Figure 1 – item 4), once that option is clicked, a configuration window will open. In here, some management actions (Figure 7.1) and a table with the different datasets sources (Figure 7 – item 2)

To add new catalogues, the user has to introduce the catalogue's URL (Figure 7 – item 3) that must be compliant with OGC standard catalogue services (CSW). To select a catalogue to be used in an evaluation, the column 'Enabled' have to be checked (figure 7 – item 4). Every time a catalogue is enabled, the platform will try to communicate through the CWS services in order to get the total number of datasets available that will be shown in the 'Total datasets found' column (Figure 7 – item 6). In the case of non-successful communication, the column *'HTML request'* (Figure 7 – item 5) may be checked to allow the use of synchronous HTML request for the communication with the catalogue server (instead of using an *Ajax* request) since some servers do not allow the default option of asynchronous requests.

Finally, in order to save the datasets' sources configuration, the user must click on the *'Save'* button (Figure 7 – item 7).

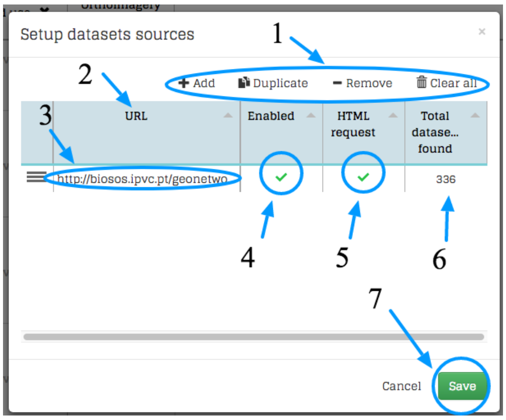


Figure 7 – Metadata database catalogues setting.

Running the evaluation

After completing the previous tasks, the evaluation process can be initiated. Each metadata record is retrieved from the selected metadata database catalogues and evaluated against user's defined requirements. The execution is started by choosing the *'Run evaluation'* option from the bottom-right menu (Figure 1 – item 4). The evaluation will start and a spinner will be shown until the evaluation ends. It should be noted that the evaluation process could be a time-consuming task when connecting to large metadata catalogue server(s). The execution time is largely dependent on the size and access speed to the metadata catalogue server(s), but also related to the Internet connection bandwidth.

Result presentation

Once the evaluation has been completed, a window with the result will open (Figure 8). The result is available in different view modes (Figure 8 – item 1), allowing to analyse which datasets are fit or unfit and to what extent they are suitable or unsuitable for a given purpose, as well as, a summary describing the criteria that are difficult to achieve.

Figure 8 – Result window | Summary view.

The summary view mode (Figure 8) presents a list of all targeted datasets identified by the respective thematic category (Figure 8 – item 2) with a summary of the fitness for use evaluation for all the datasets obtained from metadata catalogue servers for each evaluated quality indicator (Figure 8 – item 3). A legend of the symbols can be visualized using the icon on the right side of the thematic category (Figure 8 – item 4). For each targeted dataset, the list summary is divided in three groups (Figure 8 – item 5): (i) fit datasets that includes datasets which are in conformity for all expected values of quality indicators (match of 100%); (ii) partially fit datasets that incorporates datasets which are not in conformity with all quality indicators (at least one non-conformity detected) but are conformant for all quality indicators defined as critical factors; and (iii) unfit datasets which are in non-conformity with at least one quality indicator identified as a critical factor.

The *'Statistics'* view mode (Figure 9) aims to present several statistical measures about the results in order to facilitate the identification of the quality indicators that have the most influence on the success or failure of datasets availability for the application context. It presents a table with rows grouped by targeted datasets, identified by the respective thematic category (Figure 9 – item 1) and presents summary statistics for each quality indicator (Figure 9 – item 2):

- *'Total'*: the total number of evaluated metadata of datasets filtered from enabled database catalogue servers (Figure 9 – item 3);
- *'Datasets in Conformity'* → the total number (and percentage) of datasets (metadata) which are conformant with the quality indicator(Figure 9 – item 4);
- *'Excluded Datasets* → the total number (and percentage) of excluded datasets for not being in conformity for the specified quality indicator specified as a critical factor (Figure 9 – item 5).

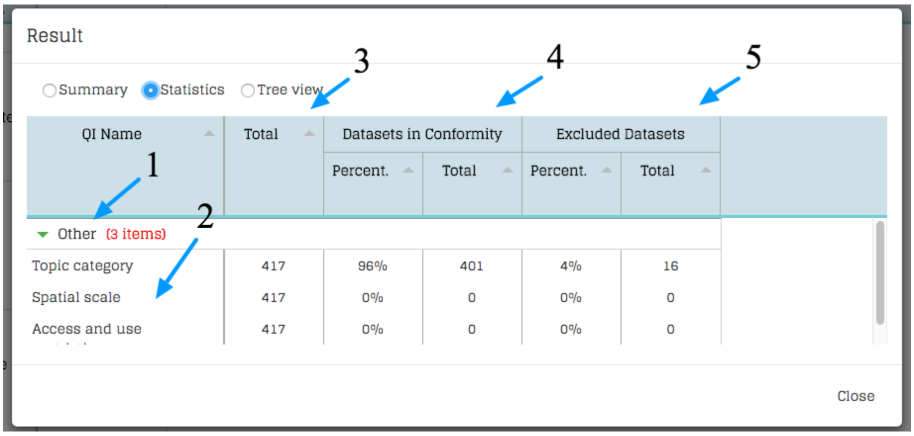


Figure 9 - Result window | Statistics view.

Finally, the *'Tree view'* viewing option (Figure 10) presents the results using a hierarchical tree-like view, which can be expanded to have a more comprehensive view of the evaluation results. The tree view is structured in the following levels:

- *Thematic Category* – the first level divides the results by the targeted datasets (thematic category) configured by the user (corresponding to the grid's columns) (Figure 10 – item 1);
- *Datasets source* – this level divides the results by metadata database catalogues selected by the user as datasets sources to be evaluated (Figure 10 – item 2), being presented the identification of the CWS URL and the total of datasets evaluated and, on the right of the row, the total of datasets evaluated (Figure 10 – item 3);
- *Dataset* – the third level contains a list of all the datasets retrieved from the catalogue of the previous tree level, being displayed the dataset title (Figure 10 – item 4), on the right of the row, the global fitness for use value for the defined quality indicators and the global fitness value for the metadata/ quality evaluation (Figure 10 – item 5);
- *Requested Metadata Evaluation* – this level is referent to the external metadata evaluation (Figure 10 – item 6) with the value of its global fitness for use (Figure 10 – item 7). The sub-levels of this group allow to see the results of conformity and critical factors of each quality indicator resulting from the comparison of user defined expected value(s) and metadata element value(s) of the dataset (third tree level);
- *Metadata Element Evaluation* – this level presents the evaluation results (Figure 10 – item 8) of the compliance of the dataset (identified on third level) relative to Metadata/Quality Elements for each configured metadata standard profiles, and on the right, a global fitness' value average (Figure 10 – item 9). The sub-levels are divided by the metadata standards profile evaluated (Figure 10 – item 10) and present, for each configured metadata/quality element (Figure 10 – item 11), the corresponding evaluation result *'In conformity'* (Figure 10 – item 12).

Besides the results' visualization modes described, the platform allows to view the metadata contents of each evaluated dataset obtained from catalogue servers (Figure 11) by clicking on the dataset title on the summary and tree view modes (e.g. Figure 10 – item 13).

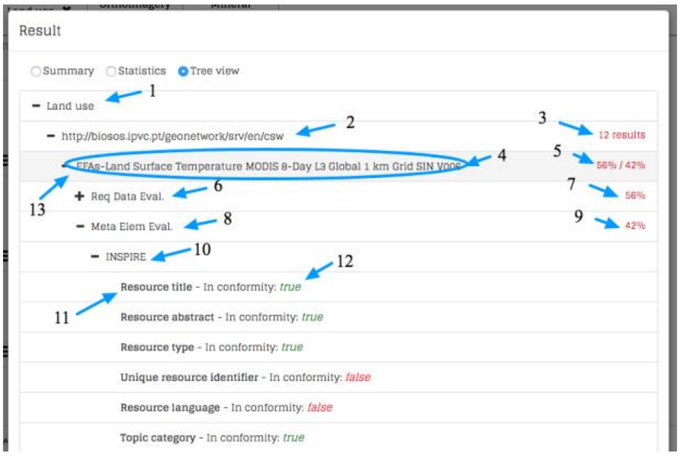


Figure 10 - Result window | Tree view.

Figure 11 - Metadata content preview.

Store configuration

All the evaluation configuration can be managed using the options available on the bottom-right menu (Figure 1 – item 4).

Figure 12 shows all the available options, from bottom to top:

- *'Save current status'*: Saves all the current configurations done by the user. This data will be removed when the user clears his browser's data or when uses the *'Clear saved status'* option of this same menu;
- *'Clear saved status'*: Clear all the configurations done and saved by the user. This will delete all the data saved by the *'Save current status'* option of this same menu, having the same effect as the clear button;
- *'Download current status'*: *Download* all the configurations done by the user to the user's device to a local file (equivalent to save the current configuration). This file can be transmitted from user to

user and used by different people to share an evaluation's configuration by using the *'Upload status'* option of this same menu;

- *'Upload status'*: Upload some configurations' file which was previously downloaded using the option '*Download current status'* of this same menu.

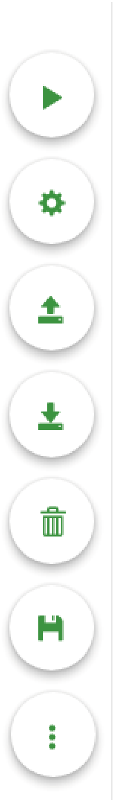


Figure 12 - Options menu.