OPTAIN

Optimal Strategies to Retain Water and Nutrients

D6.1: Common working environment with standardised metadata for the harmonised reporting of project outputs

Authors: Natalja Čerkasova (KU), Rasa Idzelytė (KU), Primož Banovec (UL), Matjaž Glavan (UL)

Delivery Date: 31. August 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No. 862756.

Disclaimer

This document reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Intellectual Property Rights

Project Consortium

Document Information

Program	EU Horizon 2020 Research and Innovation Action H2020-EU.3.2.1.1 (SFS-23-2019)
Grant agreement No.	862756
Project acronym	OPTAIN
Project full name	Optimal strategies to retain and re-use water and nutrients in small agricultural catchments across different soil-climatic regions in Europe
Start of the project	September 2020
Duration	60 months
Project coordination	Prof. Dr. Martin Volk Helmholtz-Centre for Environmental Research GmbH - UFZ www.optain.eu
Deliverable	D6.1: Common working environment with standardised metadata for the harmonised reporting of project outputs Description of the methods and tools for the harmonised reporting of project outputs
Work package	WP6: Synthesis and policy recommendation
Task	Task 6.1: Establishment of the common working environment (CWE)
Lead beneficiary	Klaipėda University
Author(s)	Natalja Čerkasova (KU, partner no. 11), Rasa Idzelytė (KU, partner no. 11), Primož Banovec (UL, partner no. 8), Matjaž Glavan (UL, partner no. 8)
Contributor(s)	Brigitta Szabó (Tóth) (ATK), János Mészáros (ATK), Felix Witing (UFZ)
Quality check	Felix Witing (UFZ), Martin Volk (UFZ)
Planned delivery date	Month 12 (August 2021)
Actual delivery date	31/08/2021
Citation	Čerkasova, N., Idzelytė, R., Banovec, P. & Glavan, M. (2021): *Common working environment with standardised metadata for the harmonised reporting of project outputs*. Deliverable D6.1 EU Horizon 2020 OPTAIN Project, Grant agreement No. 862756
Dissemination level*	PU

*PU = Public; PP = Restricted to other program participants (including the Commission Services; CO = Confidential, only for members of the consortium (including the Commission Services).

Deliverable status

Version	Date	Author(s)/Contributor(s)	Notes
0.4	13.08.2021	Natalja Čerkasova (KU)	First draft
0.5	15.08.2021	Matjaž Glavan (UL)	Validation
0.6	16.08.2021	Natalja Čerkasova (KU)	First complete draft
0.7	23.08.2021	Natalja Čerkasova, Rasa Idzelytė (KU)	Revision
0.8	24.08.2021	Matjaž Glavan (UL)	Validation of revision
0.9	28.08.2021	Martin Volk (UFZ), Felix Witing (UFZ)	Quality check
1.0	31.08.2021	Natalja Čerkasova (KU)	Final

Summary

We analysed standards for the common working environment (CWE) and established a platform for project partners' harmonised reporting of project outputs. The results of OPTAINs work packages addressing specific issues of the project (actor platform, measures and indicators, data mining, modelling, optimisation) are assimilated in the following way:

- Basic project integration is handled through the coordination of all elements of the project. These elements include the project management structure as defined by OPTAIN's Grant Agreement and Consortium Agreement, Data Management Plan (D8.1), Knowledge Management Plan (MS24), Communication and dissemination strategy (D7.2), and data sharing standards.
- Cross WP and cross catchment analysis are enabled by harmonising tools, data, and process workflows. Following the ISO 21500:2012 Guidance on Project Management, the Icam DEFinition for Function Modeling (IDEF0) is used as a tool to enable such harmonisation, where database will provide possibilities for cross-Case Study analysis.
- Exchange of modelling and communication with the public will be coordinated by following the developed protocols, metadata standards via the data-sharing platforms. Selected exchange tools and platforms follow the EU rules on open access to all scientific peer-reviewed publications and research data (Article 29.2 of the Model Grant Agreement) as well as the FAIR (Findable, Accessible, Interoperable and Reusable) data principles. In addition, custom tools were created to semi-automate the generation of metadata files to ease project data and result dissemination.
- Providing the information base for the OPTAIN Learning Environment is enabled by ensuring that the OPTAIN project consortium utilises the established CWE platform and stores the data (or information) in a joint project database. Since the data, information, resources, and project results are harmonised across all the case studies and work packages, the Learning Environment can utilise these tools or use any database views to produce the necessary data or information. For example, Learning Platforms could utilise Application Programming Interfaces (APIs) to extract the needed data in the required format.

The established CWE provides an understanding of the technical and functional background of the knowledge produced in the project and enables communication between project partners and stakeholders at different management levels and provides easy sharing of project results and recommendations.

From a technical standpoint, the CWE will bring together data that is shared among partners using a cloud and/or Structured Query Language (SQL) server. Version control of the created tools and models will be handled by utilising the Git software. Data sharing with the public will be performed through Zenodo, a general-purpose open-access repository developed under the European OpenAIRE. Metadata standards will be ensured by creating the metadata Extensible Markup Language (XML) files for every produced geospatial dataset.

Table of Contents

Abbreviations

API	Application Programming Interface
CCT	Consortium Coordination Team
CDS	Communication and Dissemination Strategy
CS	Case Study
CSS	Case Study Sites
CWE	Common Working Environment
DMP	Data Management Plan
DOI	Digital Object Identifier
EER	Enhanced Entity–Relationship
FAIR	Findable, Accessible, Interoperable and Reusable
IDEF	Icam DEFinition for Function Modeling
MARG	Multi-Actor Reference Group
NSWRM	Natural Small Water Retention Measure
ORDP	Open Research Data Pilot
SQL	Structured Query Language
WP	Work Package
XML	Extensible Markup Language

1. Introduction

1.1. Purpose of the Common Working Environment

The OPTAIN project aims at a harmonisation of its approaches across all 14 case study sites and biogeographical regions. Standardised processes and reporting as well as data harmonization are implemented to combine data from different sources, to provide project partners with comparable data from different case studies, and to allow for a strong synthesis and regional comparison of OPTAINs results.

The Common Working Environment (CWE) provides an understanding of the technical and functional background of the knowledge produced in the project. Specific information (stakeholder platform, measures and indicators, physical data on catchment characteristics, integrated modelling, optimisation of solutions, policy arrangements), required by Work Packages (WPs) and gathered on the Case Study Site (CSS) level (horizontally), will be linked across WPs (vertically) (Figure 1). Essential results of WPs 1-5 will be integrated into a CWE platform. The platform allows project partners to extract the preferred information that can be used or linked for further work on the project. Moreover, it allows WP6 (project synthesis & policy recommendations) and WP7 (learning environment) to efficiently extract, compare, integrate, and present information on the level of CSS or biogeographical regions.

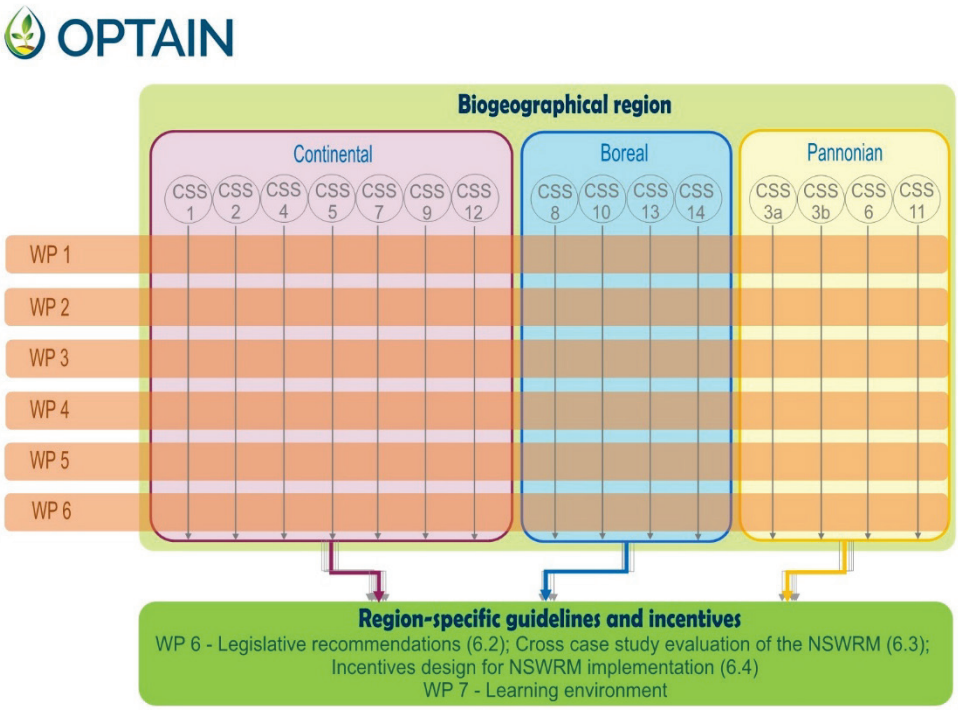


Figure 1: Schematic representation of a common working environment (CWE) designed in WP6.

The CWE platform will reveal its full potential in task 6.3 (Cross case study evaluation of the physical and economic effectiveness of the different individual NSWRM and their combinations). The data collected in the platform will be used to compare and evaluate the physical, environmental, and economic effectiveness of individual Natural Small Water Retention Measures (NSWRMs). Further, the effects of their combination and allocation across various spatial conditions in different European biogeographic regions

and agricultural systems will be studied. Region-specific guidelines for the optimal combination of NSWRM and their implementation will be designed based on inputs and outputs of WP1-WP6. The CWE will also support task 6.4 (Development of incentives for promotion of techniques of water and nutrient management based on Sustainable Development Goals) by the development of incentives promoting NSWRM.

1.2. Related Deliverables and Milestones

The Horizon 2020 project OPTAIN takes part in the Open Research Data Pilot (ORDP) of the European Commission (EC). The ORDP aims at enabling and improving access and reuse of research data generated by H2020 projects. Being part of the ORDP requires developing and updating a Data Management Plan (DMP) for the OPTAIN project.

Deliverable 8.1: Data Management Plan (DMP). The CWE deliverable (D6.1) is tightly related to the Data Management deliverable (D8.1); hence, these documents complement each other. The DMP defines concrete measures by which data is stored, managed, protected, and resources are allocated. The CWE presents a broader analysis of standards for the joint project implementation domain and provides suggestions on the establishment of a common platform. The main difference is that the DMP sets the data provision to external users, and the CWE sets the standard of the project internal data processing. The CWE was established during the first year of project implementation, whereas the DMP will be updated periodically during the entire project period.

Milestone 13: Analysis of standards for the common working environment. The MS13 preceded this deliverable, where a broad overview of common working environments and tools was shared within the OPTAIN consortium. During the screening, some tools were tested or assessed for their suitability and usability. Here we present only the agreed-upon standards of the OPTAIN project working environment and describe the tools used to facilitate this process.

Milestone 24: Knowledge Management Plan (KMP). The Knowledge Management Plan presents and analyses general knowledge pathways into, within, and out of the OPTAIN project. It documents how the implementing partners bring knowledge into the project (from research, from a large variety of actors, from other projects), how to share information, progress, and data within OPTAIN during the project lifetime; and how to communicate, disseminate and exploit the results and outputs of OPTAIN during and after the project lifetime. Similar to the DMP, the KMP serves as an overarching guide for all project partners as well as provides an important basis for the general project management (WP8), whereas the CWE defines concrete standards that should be followed, for joint project implementation.

2. Project integration management

For the OPTAIN project integration management, we adapted a systematic approach to coordinate all the processes and activities of a project. We treated the entire project as a complex system, where tasks and associated deliverables are treated as processes, and information or data flows between them as links. We used several tools to analyse and harmonise each system entity (Figure 2).

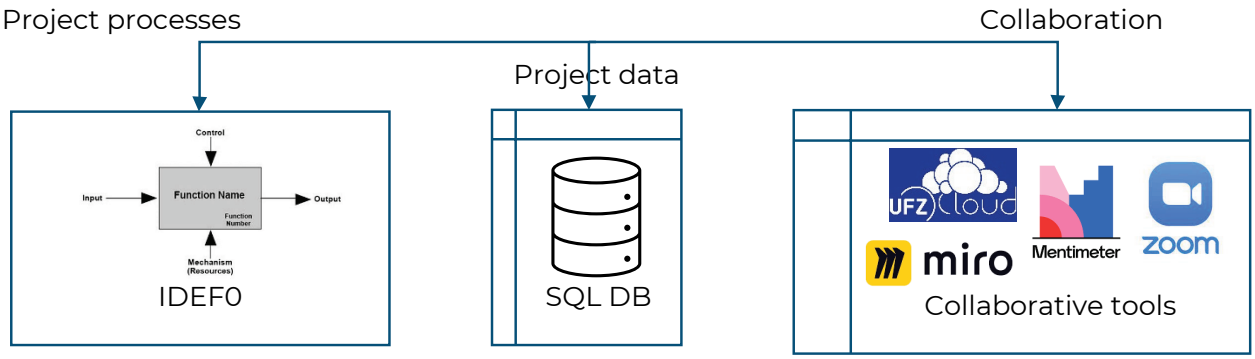


Figure 2. Project integration conceptual scheme with tools

2.1. Project process harmonisation

Project process harmonisation was focused on designing, agreeing on, and implementing process standards across different work packages to facilitate achieving targeted project results while ensuring a harmonious acceptance of the new processes and information/data flows by the different actors. A conceptual approach was adopted to identify the data requirements and information exchange. The scoping was performed in two stages:

a) Current data collection - general stage (general description of expected inputs, applied working processes in specific WP Tasks (modelling, communication etc.) and produced outputs (Deliverables, Milestones). In addition, a D6.1 questionnaire on the identification of the data requirements and information exchange was sent out to all task leads, asking them to fill in a form (presented in Appendix 1). The answers were systemised in the DMP (refer to Tables 1-7: "Listing and short description of the outputs and datasets" in the DMP document) and in the CWE Milestone report.

b) Detailed stage – second screening and validation of the processes, information flows and outputs. A project partner panel revised the questionnaires, systemised the information and the data flows, identified the issues in the process flows and returned the questionnaires to the WP and Task leads for correction and validation.

The resulting project process flow model (described in the section below) is a maintained document. New processes and links will be included and updated, following updates in the DMP and other related documents.

2.1.1. Process Functional Modelling

Icam DEFinition for Function Modeling (IDEF0) is a technique for the structured analysis and design of systems. Its use in improving productivity and communications in manufacturing systems and as a tool for business process reengineering efforts are

widely documented (Presley and Liles, 1995). While the original use of IDEF0 notation was intended for military engineering purposes, the method is designed to model any system's decisions, actions, and activities, no matter the size, complexity, or target result.

We adapted the IDEF0 technique to model the internal processes of the OPTAIN project. Each work-package deliverable or relevant milestone is considered a process. To enable any process (deliverable), a specific input is required. One or more controlling mechanisms use available resources to process the inputs and produce an output (result). The output of any process is used in other internal processes or is exported to an external system via project dissemination, outreach channels or other methods, detailed in the dissemination strategy (WP7). The simplified explanation of IDEF0 notation is presented in Figure 3.

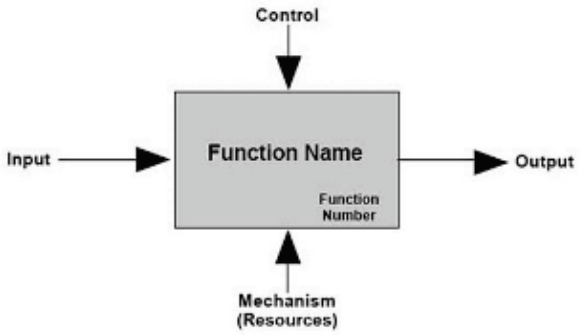


Figure 3. IDEF0 notation explanation

The complex cross-WP and cross-task interactions can be analysed in detail using this notation. Figure 4 presents the entire OPTAIN system/project.

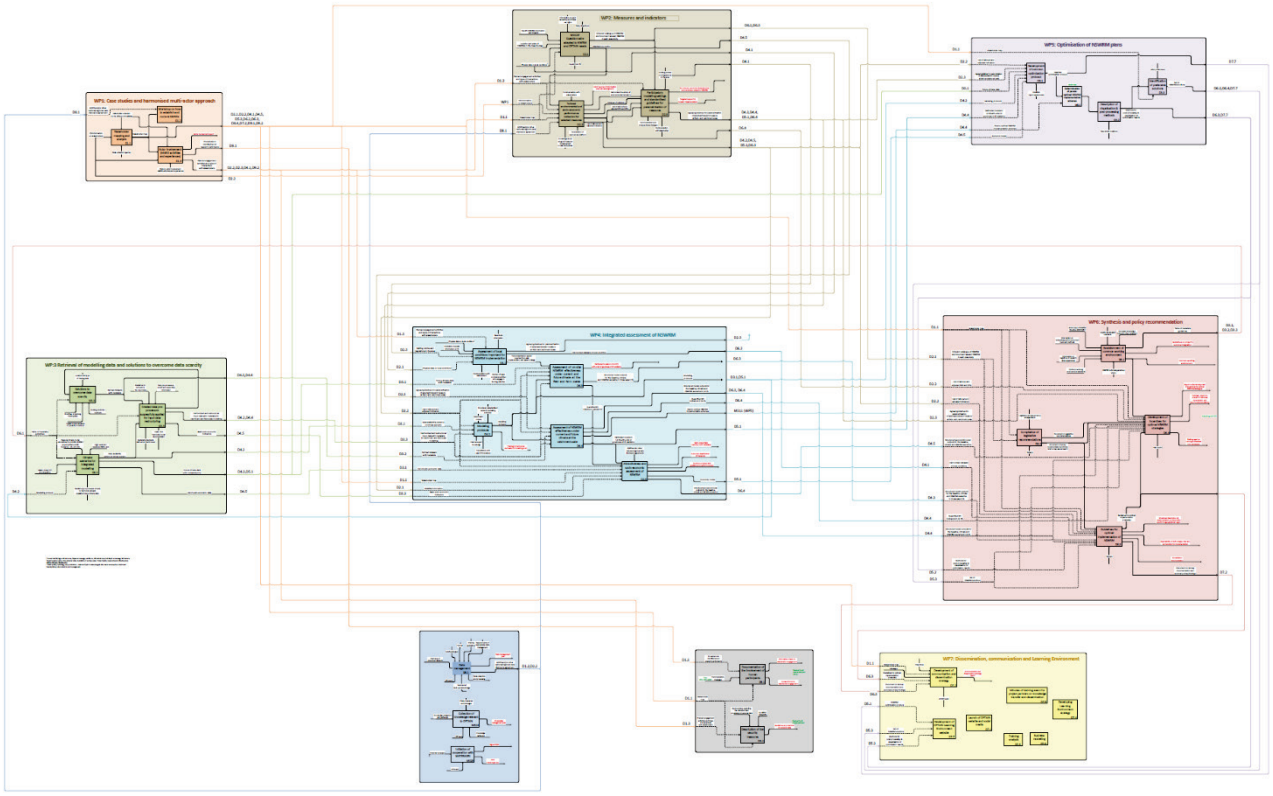


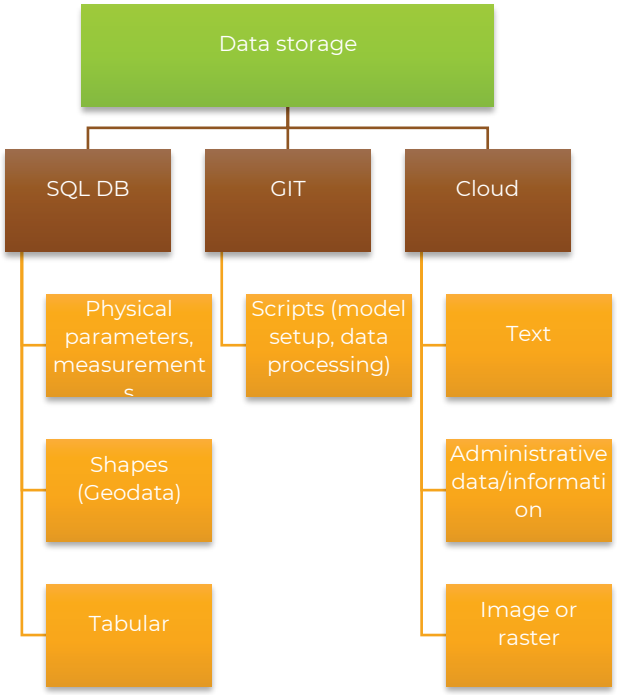
Figure 4. IDEF0 diagram of the OPTAIN project

Due to the complexity of the modelled system (OPTAIN project), dedicated IDEF0 diagrams for each WP can be found in Appendix 2.

This representation of the project internal structure and information streams helped WP and task leads identify the disagreements (structural or contextual) between the project processes. These disagreements were addressed internally and resulted in a common understanding of the technical and functional background of the data, information and knowledge produced in the project. Despite the differences of the case study sites, each process related to the activities in these sites will be carried out in a harmonised way based on defined and agreed-upon protocols, producing intercomparable data and information.

2.2. Project data harmonisation

During the OPTAIN project implementation, many kinds of data will be collected, processed, and stored (various data types and formats are described in detail in the DMP). In order to ensure data harmonisation, a certain standardised approach must be implemented and followed. Furthermore, due to the different nature of data, different platforms will be used to accommodate different types of structured information



(Figure 5).

Figure 5. Data storage solutions for various types of data produced or accumulated during the project

SQL Database. Structured Query Language (SQL) is a domain-specific language used to manage data held in a relational database management system and handle structured data, i.e., data incorporating relations among entities and variables (Beaulieu, 2009). The SQL based database was designed for data exchange and storage within the project, e.g., to make data on physical characteristics of all case study sites easily accessible for modelling by any partner modeller. The main advantages of a relationship database are:

i) Data integrity. SQL database allows setting the required (mandatory) fields, without which the datasets cannot be entered.
ii) Increased consistency and fewer updating errors. Each field in the database can have a set of either default values or value boundaries. Moreover, each field has a specific type of data, e.g. it is impossible to write text or any random number to a field, which is defined as DATETIME format.
iii) Defined data relations. Data relations are defined by the database scheme, which can also be visualised for clarity of the overall data concept and communication of responsibilities to users in their different roles. If the need arises, the scheme may be changed to allow other or new functionalities.
iv) Independence from applications or programs. SQL databases can be connected through various tools for all common operating systems and dedicated libraries for programming languages, such as Python, R, and others.
v) Data security. Access to the databases is secured through password-protected users, preventing unauthorised access (providing either read and/or write access rights).

These advantages also assure that the research data meets the FAIR principles - findable, accessible, interoperable, and reusable.

GitLab. GitLab is a web-based product lifecycle tool that provides a Git repository and managerial solutions, such as issue-tracking, continuous integration, and deployment pipeline. GitLab is an open-source project. In the OPTAIN environment, the most important features are version control and the continuous integration of common project software and scripts. The UFZ GitLab service (found at https://git.ufz.de/) is currently used for hosting and editing code, managing project issue-board, and building software.

UFZ Cloud. Cloud storage (referred to as OPTAIN cloud) was established to store and backup all project-related data and documents. The OPTAIN Cloud is hosted on UFZ (Helmholtz Centre for Environmental Research) servers in Leipzig (Germany) and uses the open-source software Nextcloud.

The document naming, formatting and versioning protocol was agreed upon and is detailed in the Data Management Plan. Following the ISO 21500:2012 Guidance on Project Management and ISO 9001:2015 Quality Management System Requirements, regular snapshots (backups) of the entire Cloud are made. The cloud storage has an intuitive structure (separate folders for administration purposes, WP-related work, and each Case study), which was agreed upon early in the project. Every partner received an email familiarising themselves with this platform, usage, and login credentials. All OPTAIN partners have equal access rights to the Cloud, which is subject to change in the future.

2.2.1. Database design

An iterative waterfall model was chosen when designing the relationship database, as this approach provides feedback paths from every phase to its preceding phases (Figure 6).

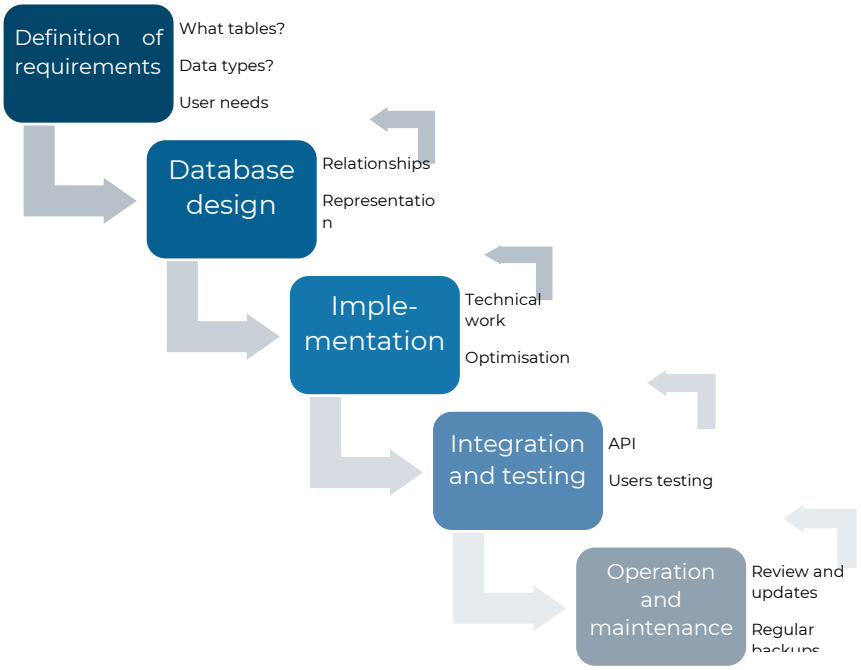


Figure 6. The iterative waterfall design process of the SQL DB for the OPTAIN project

Firstly, specific needs for data were identified, followed by the determination of data types. This process was interlinked with the OPTAIN DMP, where the foreseen information and data types are described in detail. Next, specific relationships were identified, and their multiplicity was determined. Finally, an enhanced entity-relationship (EER) model was used to coordinate the requirements of the users (project implementers) with technical possibilities (SQL data formats and design standards). The resulting model is shown in Figure 8.

The notation of the EER diagram is standardised (R. Elmasri, S. B. Navathe (2011). The main elements are 1) entities, 2) attributes, 3) relationships. In the OPTAIN database, the entities are represented by objects, which contain attributes. Objects are connected via relationships, which represent multiplicity. An example of a simple database EER model is given in Figure 7. The two entities "user" and "category" are linked via a "one-to-many" relationship, meaning that a "user" may be only of a single "category", but many "users" may have the same or different "categories".

Given the current state of development the CWE platform database will store information that can be broadly categorised in such classes:

- *Measurements*: containing data on physical measurements, such as water quality samples, timeseries, soil samples, crop production reports, etc.
- *Models*: data on model setups for each catchment/case study, model characteristics and links to used datasets.
- *Input data*: processed or raw datasets needed to initiate any model setup, such as landcover data, topographic data, etc.
- *NSWRM*: information on individual measures, which will be considered and used for the optimisation scenarios.
- *Results*: scenario outputs of different models and condensed information on the interpretation.

- *Supporting tables*: information needed for harmonisation and referencing, i.e. country, case study, modelling tool, etc.
- *Views*: database views, which provide structured information from multiple tables according to user-defined criteria.

SQL database can store and interlink different data types using unique keys (database entry identifiers). We foresee that most of the data will be user-oriented, meaning that any information which is stored, will be useful to produce the database views. Because the database implementation is designed in an iterative approach, we will remove redundant information from the database with every integration and testing iteration (Figure 6). This CWE deliverable report will not describe the EER notation in exhaustive detail. However, this information is available in multiple chapters of Software Engineering books and guides, such as Ramez Elmasri, Shamkant B. Navathe (2011).

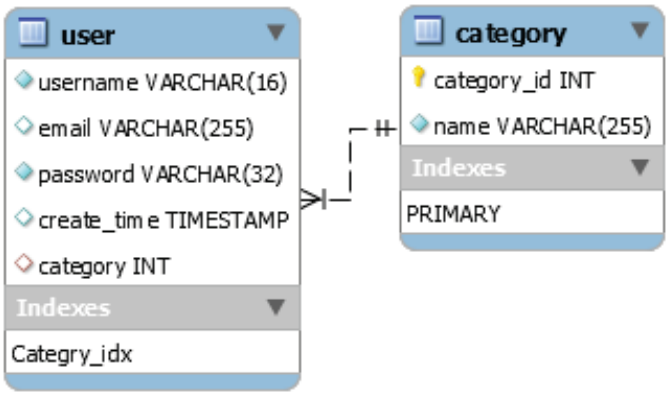


Figure 7. Example EER model of a simple database: entities are "user" and "category", attributes are "username", “email”, etc., and the “one-to-many” relationship is the interconnecting line

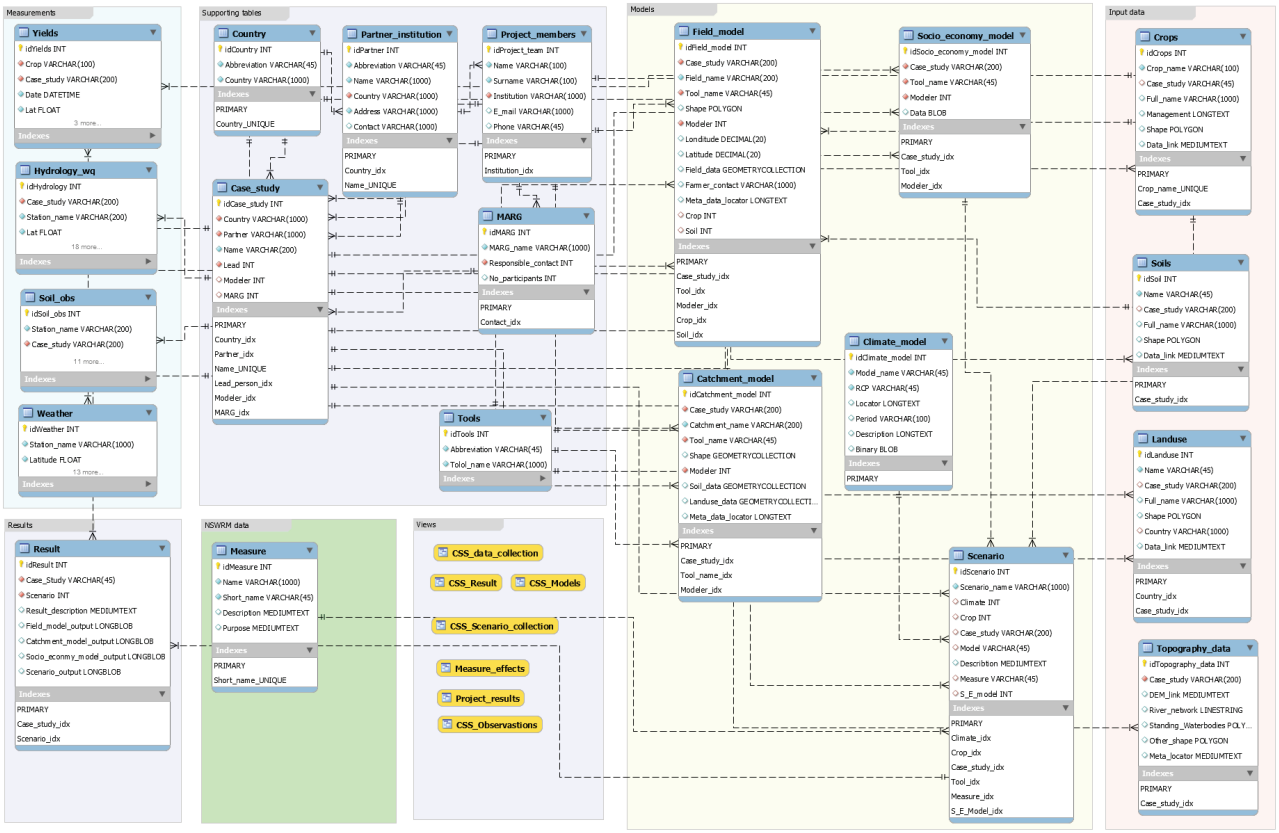


Figure 8. OPTAIN SQL database enhanced entity-relationship (EER) model

Currently, the database is in its implementation stage, with a beta version available and in the internal testing phase.

2.3. Collaborative environments

Real-time communication is necessary to accelerate the conversation about project-related workflows, deliverables, or milestones and connect with external parties, such as MARGs. The KMP (Section 4. Knowledge pathways within OPTAIN) describes the framework for the project internal communication and exchange of information, considering a variety of web-based communication formats, personal meetings and structures for the management and exchange of data. It details the OPTAINs organizational structures and all the individual communication formats that enable a sound exchange of knowledge, progress, and data in OPTAIN. Here we present some of the solutions, that were agreed-upon and are in use by the partners.

Since physical meetings sometimes can become a challenge during these pandemic influenced times, virtual communication tools, such as Zoom, Skype, MS Teams, etc., have already been tried and successfully tested approaches for communication of remote teams. In addition, virtual message boards or interactive whiteboards are good tools for sharing ideas and give feedback both within the organisation and externally.

Two collaborative environments are used for an interactive collaboration in the OPTAIN project workshops – MIRO and Mentimeter. An online collaborative whiteboarding platform MIRO (https://miro.com/), is a straightforward but comprehensive tool to fasten the thinking process. Allows dispersed working groups to meet, share ideas, and make decisions creating easy to understand flowcharts or diagrams using simple digital sticky

notes. MIRO also offers integration with additional tools, i.e. MS Teams, Zoom, GitHub, Excel, Google Drive, etc. Mentimeter is an interactive presentation software (https://www.mentimeter.com/), appealing to different actor groups due to its easy involvement and collaboration approach. It provides live polls, word clouds, multiple-choice questions, etc., to engage and get anonymous feedback from the participants of meetings or workshops, increasing efficiency and productivity of remote and on-site communications. Since Mentimeter is based on the web, it is compatible with any video conferencing platform.

3. Data exchange process

3.1. Data repositories

Projects funded or co-funded under the Horizon 2020 programme (beneficiaries of European Research Council grants) must follow certain rules on open access to all scientific peer-reviewed publications and research data (Article 29.2 of the Model Grant Agreement (H2020 Programme, 2019), following the FAIR data principles. Scientific peer-reviewed publications and all relevant data must be publicly available to read online and downloaded free of charge to meet these requirements. The open-access mandate comprises depositing publications and data in repositories and providing open access to them (H2020 Programme, 2017). We determined the appropriate dissemination and sharing platforms based on the type of information or data (Figure 9).

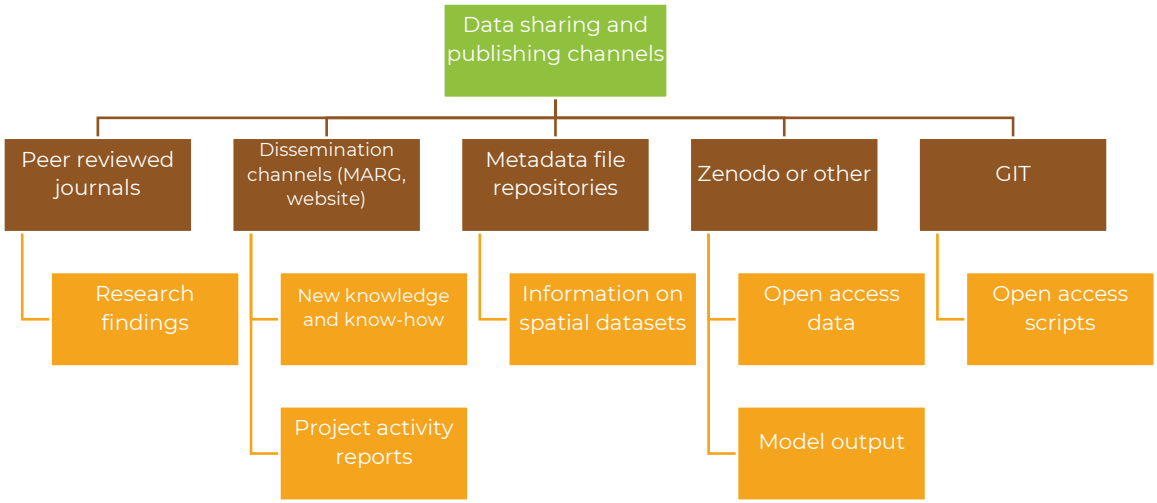


Figure 9. Information and data sharing channel conceptual scheme [1]

We suggest using **general-purpose open-access repository, such as Zenodo for storing and sharing the datasets** derived during the OPTAIN project. Zenodo is developed under the OpenAIRE program (https://www.openaire.eu/), funded by the EU, and aims to support the European Commission's Open Access policies. Zenodo allows depositing papers, datasets, software, code, reports, and other digital material related to research. While each submission receives a DOI, the data holder can choose the access type (open or closed) and share protected links with the specific viewers. Although we recommend

[1] The allocation of project outputs (orange blocks) to specific channels (brown blocks) is shown in a simplified way and specific ouput types may be shared via several channels

Zenodo, the data owner (e.g. task lead or CSS lead) has the final decision about which repository she/he wants to use, as long as its services comply with FAIR principles.

3.2. Metadata standards and tools

The INfrastructure for SPatial InfoRmation in Europe (INSPIRE) initiative (https://inspire.ec.europa.eu/) was launched in 2001 by the European Commission. The objective is "to make harmonised and quality spatial information readily available to support environmental policies or activities which may have a direct or indirect impact on the environment in Europe" (Annoni and Craglia, 2005). Commission Regulation (EC) No. 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and the Council as regards metadata sets out the requirements for the creation and maintenance of metadata for spatial data sets, spatial data set series, and spatial data services, defines the required metadata elements, their multiplicities, and the value domains to be used. In addition to these requirements, Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and the Council regarding interoperability of spatial data sets and services defined six additional metadata elements required for the interoperability.

We will create the corresponding metadata files to share the information on the datasets derived during the OPTAIN project. They will be based on the non-binding technical guidelines for implementing INSPIRE dataset and service metadata based on the ISO/TS 19139:2007 document (version 2.0.1), which are related to the regulations mentioned above. The guidelines describe all the implementation aspects together with the existing standards, technologies, and best practices. It is based on the ISO 19139:2007 standard, which provides the XML implementation schema derived from ISO 19115 (information about the identification, extent, quality, spatial and temporal schema, spatial reference, and distribution of spatial data), specifying the format for metadata records.

The metadata will be created using a simple Excel form spreadsheet. It is an easy and quick way that does not require specific user knowledge. This tool was initially developed within the GeoSmartCity project and adapted for the OPTAIN project used by the Centre for Agricultural Research, HU, and Klaipeda University, LT. In the form, some cells can be edited as free text, others contain values taken from the dropdown list, and some contain already predefined information. Furthermore, all fields have additional information in the comments window, which simplifies the user's involvement in filling out the form. Finally, the metadata files can be generated one by one or bulk using a custom R script created employing the geometa package (Blondel, 2020). The description of the tools and user manual can be found in Appendix 3.

3.3. Sharing of project results

In addition to the data exchange, the OPTAIN CWE provides the necessary **channels for the communications team to extract the needed information for project outreach**, as defined by the Communication and Dissemination Strategy and other activities of WP7. These channels may come via the already harmonised internal processes and agreed upon project implementation protocols or via technology-based solutions, such as various fit for purpose Application Programming Interfaces (API) (Figure 10).

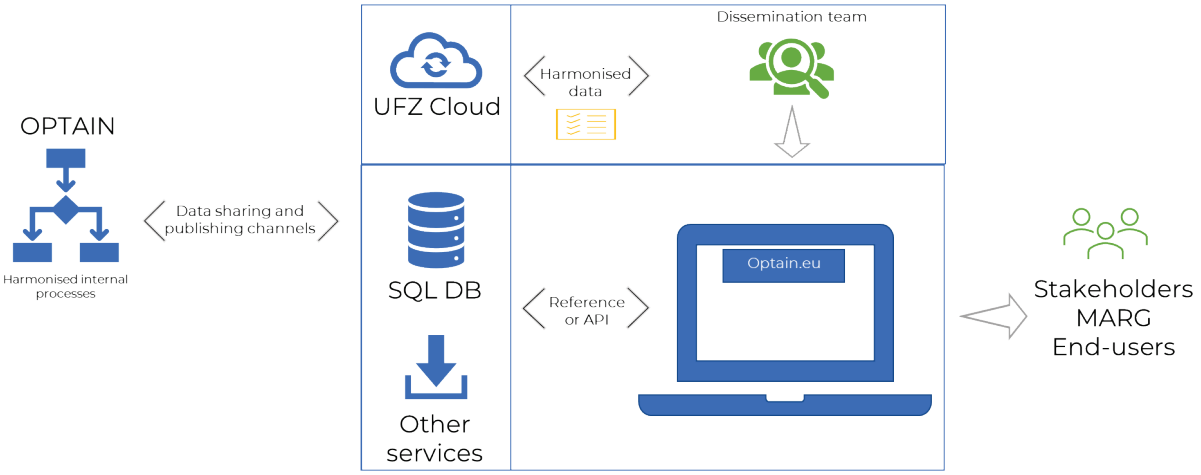


Figure 10. Conceptual scheme of the sequence for sharing and dissemination of OPTAIN data

A common scenario for such information would be a project result display on-demand on the project website or the Learning Environment. This can be achieved due to the universality of back-end solutions of the CWE: a unifying SQL database. Using general-purpose scripting language geared towards web development (i.e. PHP: Hypertext Preprocessor), the dissemination team can extract and present the data for every CSS, seeing as the stored information is harmonised and quality-controlled. This solution also ensures that the most recent project results are displayed for the stakeholder or end-user as soon as the SQL database is updated.

3.4. Linking project processes to the EU policy framework

Besides the internal OPTAIN project Input/Output structures defining the relations among the WPs, there are also external data and information exchange requirements. They are addressed as project-specific requirements that position the work of OPTAIN in relation to the external environment of specific domains. The external environment is understood as a combination of all factors outside the project that affect its processes and outcomes. However, the project itself does not affect them but instead tries to understand them and shape the results to existing settings and propose new settings to reach project aims. Examples of the external environment are political, economic, social, technological, and environmental. The domain is understood as a selected policy that will be analysed while linking project processes to achieve its objectives.

The effects of NSWRM are relevant for a wide variety of policies and simultaneously contribute to the achievement of different Sustainable Development Goals (SDGs 2, 6, 12, 13 and 15) and environmental targets formulated in several European Union (EU) policies. Many EU policy documents aiming at safeguarding and enhancing the water storage potential and foster ecosystem services for mitigating extreme weather events, such as the Common Agricultural Policy (CAP), Water Framework Directive (WFD), Floods Directive, Nitrates Directive, EU Biodiversity Strategy, EU Thematic Strategy for Soil Protection, EU Action on Water Scarcity and Drought, and the EU Strategy on adaptation to climate change (EU Adaptation Strategy).

The linking of OPTAINs project activities and procedures to the EU policy framework will be based on three main key policy domains, which are defined among the project objectives. Those domains are CAP, WFD and EU Adaptation Strategy:

A. Domain on processes of agricultural production (plant uptake, soil water retention, economic sustainability) is defined by the EU Common Agricultural Policy (CAP) with main objectives set out in article 39 of the Treaty on the Functioning of the European Union:
    i. to increase productivity by promoting technical progress and ensuring the optimum use of the factors of production, in particular, labour;
    ii. to ensure a fair standard of living for the agricultural community;
    iii. to stabilise markets;
    iv. to secure availability of supplies;
    v. to provide consumers with food at reasonable prices.
B. Domain on processes of water management (water-use efficiency, water quantity and quality) is defined in the EU Water Policy (Water Framework Directive - WFD). WFD is considered as an umbrella directive with the overall objective of at least "good environmental water body" status with main objectives:
    i. prevent deterioration and enhance the status of aquatic ecosystems, including groundwater;
    ii. promote sustainable water use;
    iii. reduce pollution;
    iv. contribute to the mitigation of floods and droughts.
C. Domain on mitigation and adaptation processes to extreme weather events (flood, drought) is defined in the EU strategy on adaptation to climate change. The strategy sets out how the EU can adapt to the impacts of climate change and become climate resilient by 2050. Climate change will impact all levels of society, and across all sectors of the economy, so adaptation actions must also be systemic. The strategy supports the further development and implementation of adaptation strategies and plans at all levels of governance with three cross-cutting priorities (integrating adaptation into macro-fiscal policy, nature-based solutions for adaptation, local adaptation action). The strategy has three principal objectives:
    i. make adaptation smarter,
    ii. make swifter and more systemic adaptation,
    iii. to step up international action on adaptation to climate change.

The achievement of a single domain-based objective is a challenge while resolving the conflicts and improving the synergies among the mentioned is a task on another level (EN, 2014) (Figure 11).

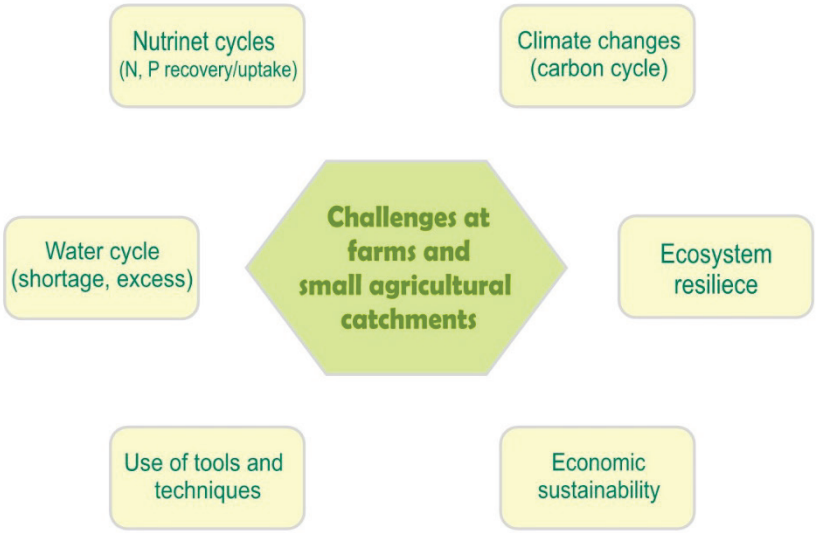


Figure 11: Challenges at multi-domain objectives

With the OPTAIN project addressing various external domains, it is important to identify their related processes, components, and units (WFD, CAP, EU Adaptation Strategy). The CWE with the definition of processes, entities and resources is required. Therefore, for the OPTAIN project, the process model is defined for domains aiming at:

(i) identification of relationships among the entities that connect analysed domains;

(ii) identification of relationships that connect domains to the processes of the OPTAIN project.

In addition, a variety of domains provide a wide range of domain-specific standards and procedures. More on the specific linkages between CAP, WFD, EU Climate Adaptation Strategy and other polices together with their impact on NSWRM will be studied in tasks 6.2, 6.3 and 6.4.

References

Annoni, A., Craglia, M., 2005. Towards a Directive Establishing an Infrastructure for Spatial Information in Europe (INSPIRE). Proceedings of the 8th International Conference on the Global Spatial Data Infrastructure (GSDI-8) - from Pharaohs to Geoinformatics p. 1-14 vol. TS 47 - SDI Legal and Policy Issues

Beaulieu, A., 2009. Learning SQL, 2nd ed. O' Reilly Media, Inc.

Blondel, E., 2020. Geometa: Tools for Reading and Writing ISO/OGC Geographic Metadata. R package version 0.6-3. https://cran.r-project.org/package=geometa

Craglia, M., 2005. Towards a Directive Establishing an Infrastructure for Spatial Information in Europe (INSPIRE). Proceedings of the 8th International Conference on the Global Spatial Data Infrastructure (GSDI-8) - from Pharaohs to Geoinformatics p. 1-14 vol. TS 47 - SDI Legal and Policy Issues.

EN, 2014. Special report – European Court of Auditors. Integration of EU water policy objectives with the CAP: a partial success (ISBN 978-92-872-0028-0 doi:10.2865/15216)

H2020 Programme, 2017. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Version 3.2.

H2020 Programme, 2019. AGA – Annotated Model Grant Agreement. Version 5.2.

Presley, A., Liles, D., 1995. Use of IDEF0 for the design and specification of methodologies.

Ramez, E., Shamkant B. N., 2011. Fundamentals of Database Systems (6th ed.). Pearson/Addison Wesley. Chapters 8 and 9. ISBN 978-0-136-08620-8.

Appendix

Appendix 1. D6.1 Questionnaire template

WP6.1: Questionnaire: identification of the data requirements and information exchange

Prepared by: add organisation acronym	DX.X* - Add deliverable number and name		
	Description of the inputs of the WP deliverable and associated milestones elaborations *(indicate WP used for inputs – if applicable)*	Description of the elaborations performed *(models, communication, queries, data collection, literature review…)*	Description of the outputs of the WP deliverable and associated milestones *(indicate also the receiving WP)*
	INPUTS	ELABORATIONS	OUTPUTS
Spatial entity information including indicative resolution and potential sources of information (cadastre, topology LIDAR, point data, INSPIRE spatial entities)	Please add text… **Indicate** from which WP and deliverable the inputs are expected if not obtained within the deliverable	Please add text…	Please add text… **Indicate** to which WP and deliverable the outputs will be provided (if provided for any other WP and deliverable)
Physical data related to the spatial entity - (soil profile data, constructive measures, non-constructive measures climate data, temporal and spatial resolution)	Please add text…	Please add text… Describe the elaborations, performed within the activities related to the specific deliverable, i.e. data collection, modelling, validation, optimization, communication with stakeholders, programming, … Usually the elaborations integrate the contents from multiple inputs and provide results for multiple outputs.	Please add text…
Inputs on legal or physical persons -and relations among them (governance issues) also potentially subject to GDPR – description of the data sources and attributes associated	Please add text…		Please add text…
Legal constrains – description of enabling environment – which sector, decrees as well,	Please add text…		Please add text…
Socio-economic data - farm accounting, social structure of farms, gross margin, subsidies….	Please add text…		Please add text…
Other – please add if needed	Please add text…	Please add text…	Please add text…

**Example description: D7.4 - Learning environment development strategy*

Appendix 2. Project process IDEF0 diagram for each WP*

**as identified at the first stage of data management questionnaire*

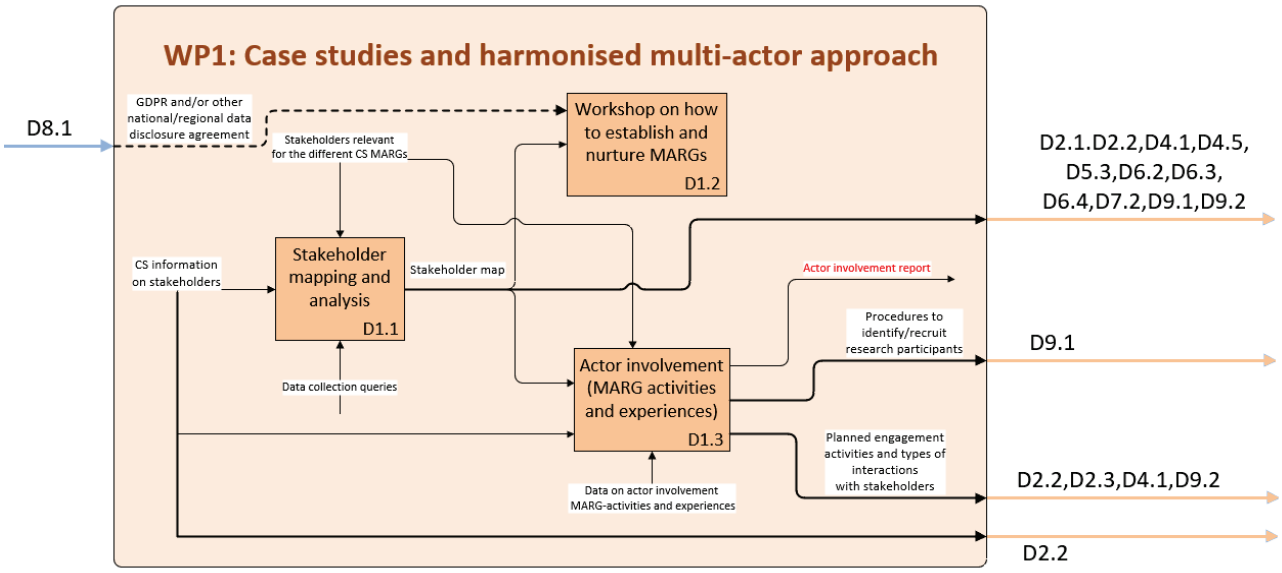


Figure A2.1. WP1 IDEF0 diagram

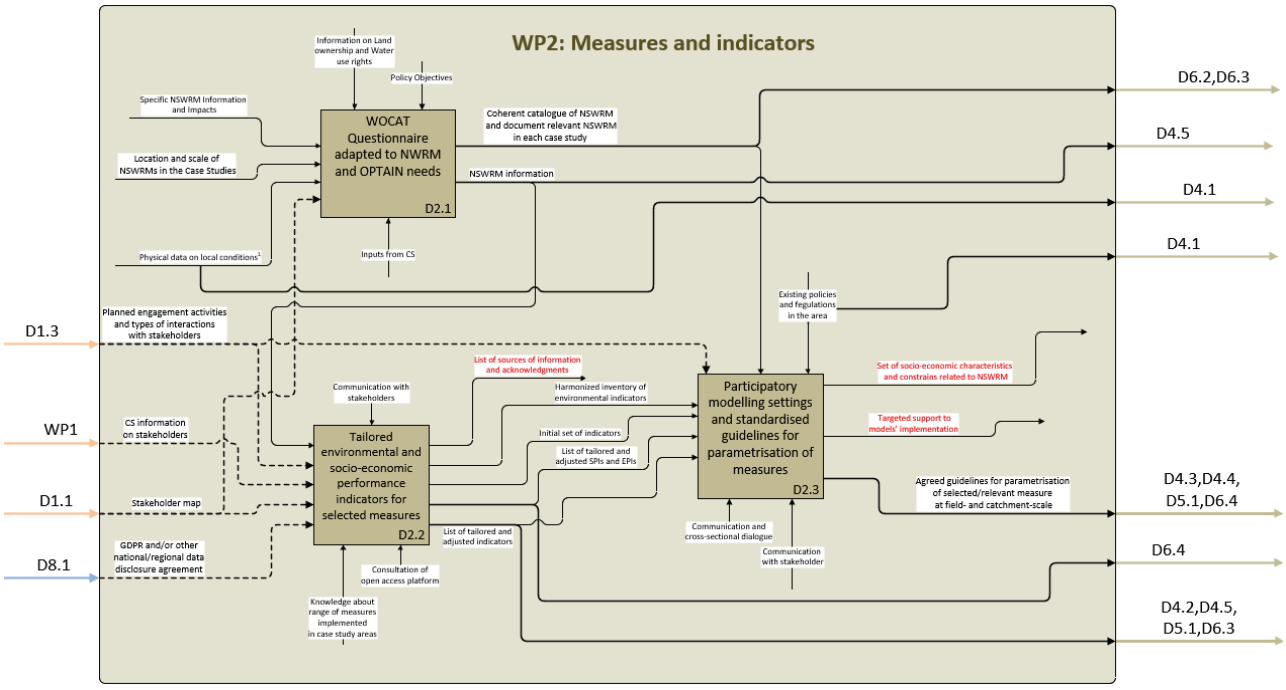


Figure A2.2. WP2 IDEF0 diagram

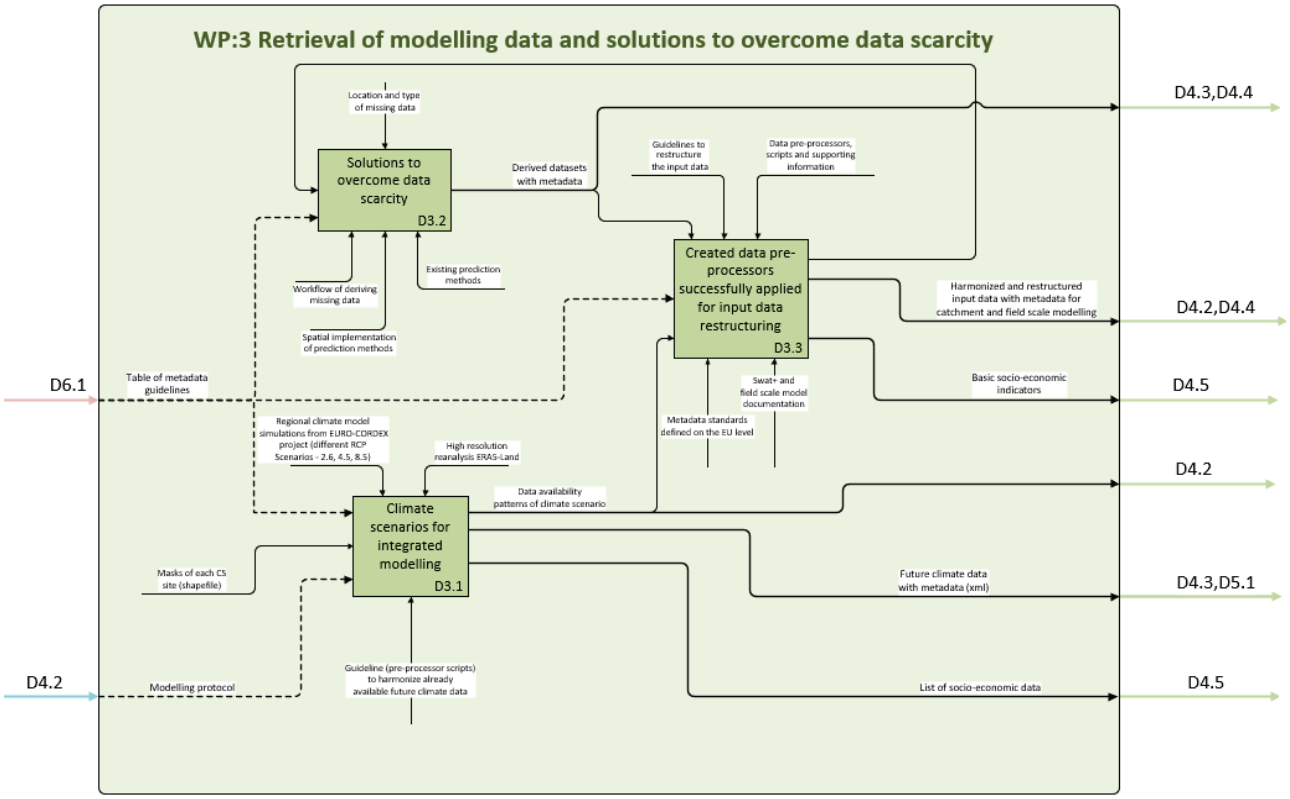


Figure A2.3. WP3 IDEF0 diagram

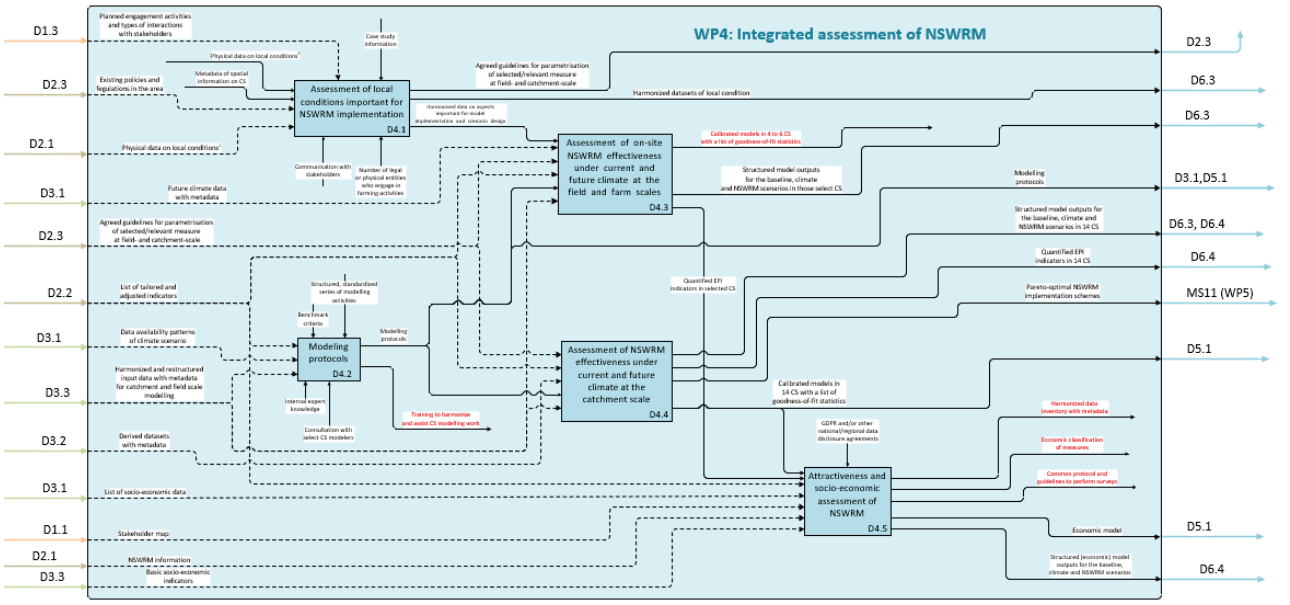


Figure A2.4. WP4 IDEF0 diagram

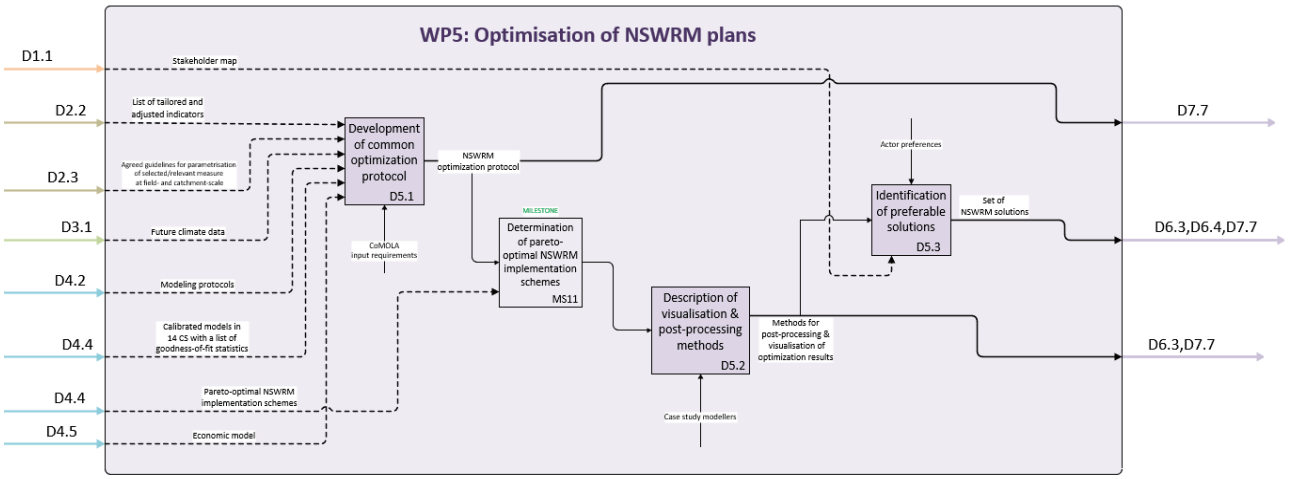


Figure A2.5. WP5 IDEF0 diagram

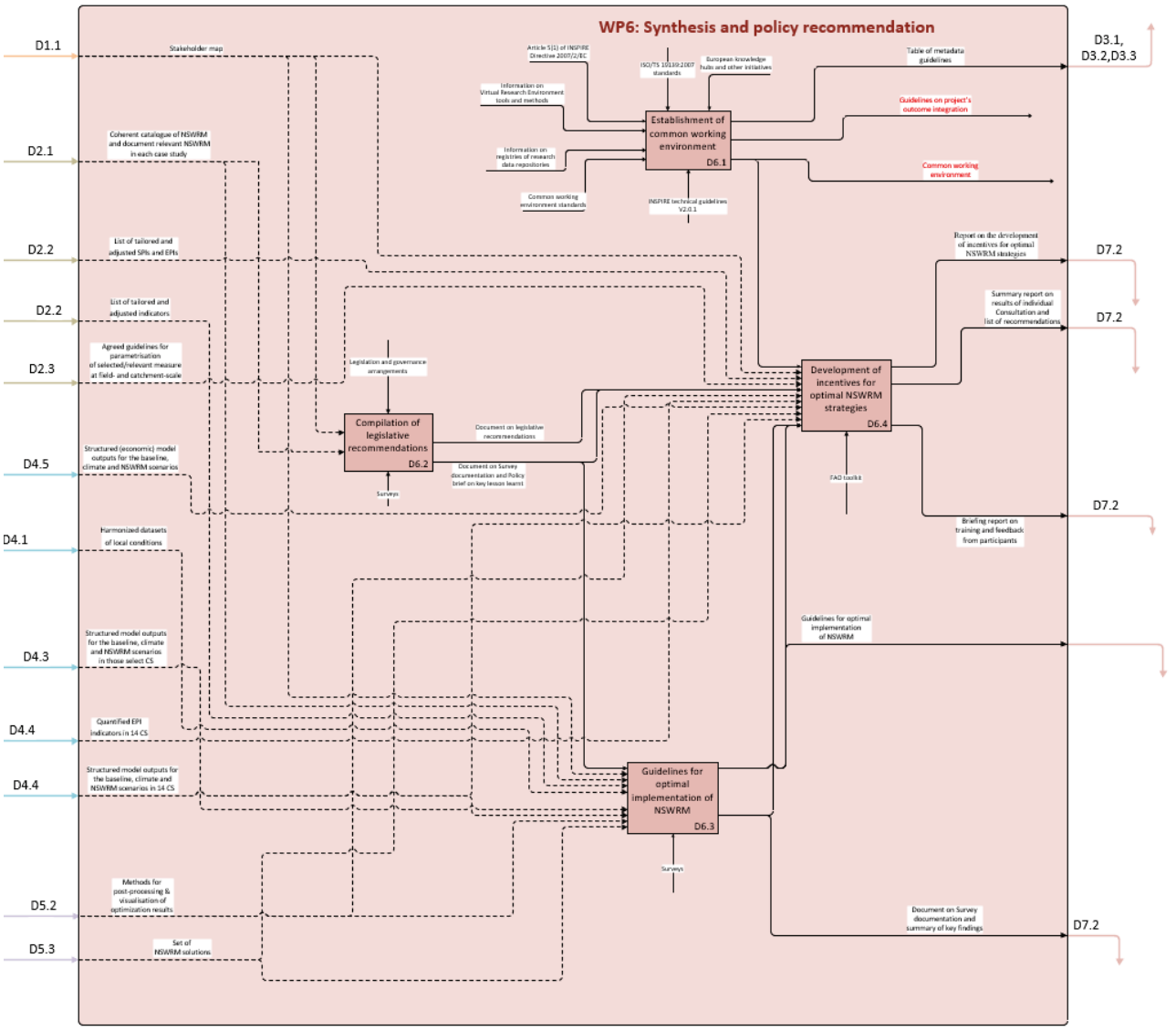


Figure A2.6. WP6 IDEF0 diagram

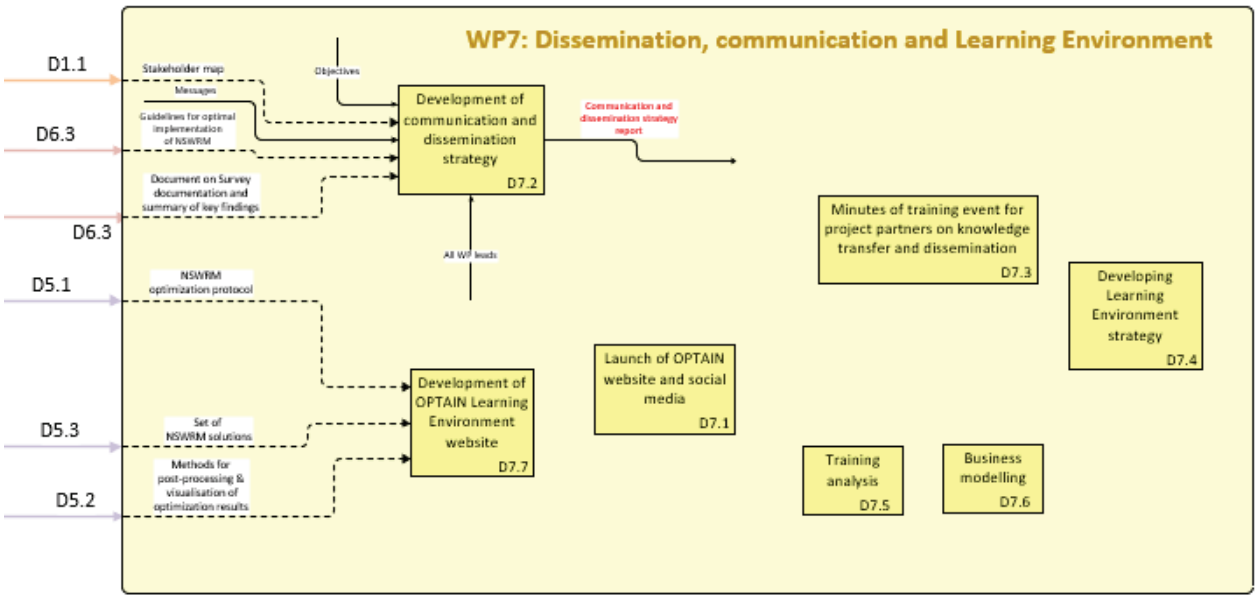


Figure A2.7. WP7 IDEF0 diagram

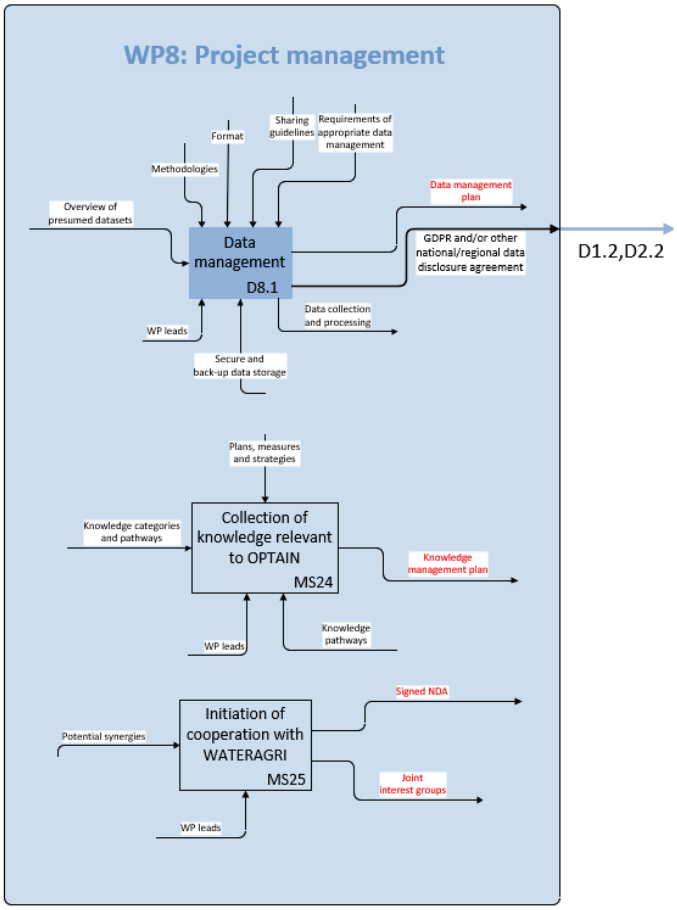


Figure A2.8. WP8 IDEF0 diagram

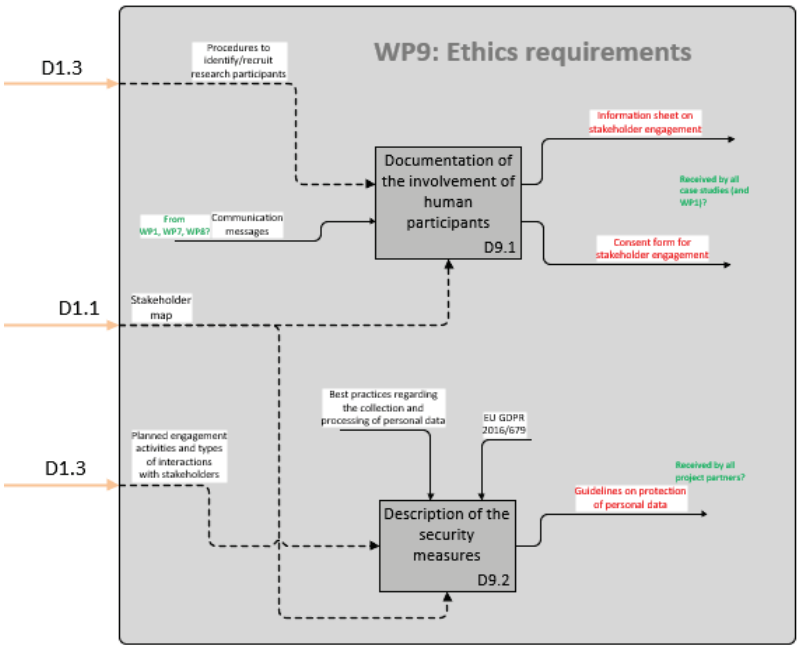


Figure A2.9. WP9 IDEF0 diagram

Appendix 3. Metadata creation tool documentation

We prepared a simple tool to collect INSPIRE XML19139 compliant metadata of spatial datasets. The complete manual and the tools are available in the OPTAIN Cloud: WPs & Tasks\WP3\Task_3_1\templates\Metadata_script\

→ Metadata_form.xlsm

→ metadata_creator.R

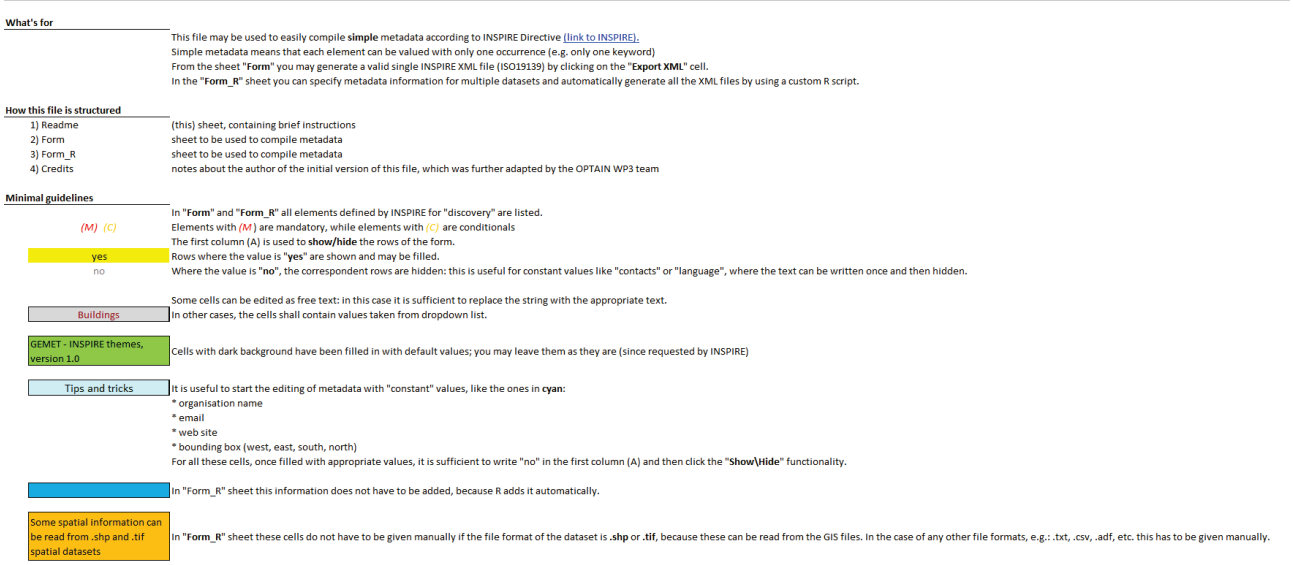
→ Metadata_user_manual2.docx

→ Metadata_tables_updated.xlsx

In the provided Excel file (Metadata_form.xlsm) the user can find four sheets (Readme, Form, Form_R and Credits). Their functioning is explained below:

I. *Readme*

This sheet provides general information about the purpose and structure of the file with some basic guidelines for filling up the metadata form. The gray colored cells in the further form sheets indicate that values have to be chosen from a predefined drop-down list, while green colored cells shall not be edited as they contain the preset information from the INSPIRE directive technical guidelines.



What's for

This file may be used to easily compile **simple** metadata according to INSPIRE Directive (link to INSPIRE).
Simple metadata means that each element can be valued with only one occurrence (e.g. only one keyword)
From the sheet "**Form**" you may generate a valid single INSPIRE XML file (ISO19139) by clicking on the "**Export XML**" cell.
In the "**Form_R**" sheet you can specify metadata information for multiple datasets and automatically generate all the XML files by using a custom R script.

How this file is structured

1) Readme	(this) sheet, containing brief instructions
2) Form	sheet to be used to compile metadata
3) Form_R	sheet to be used to compile metadata
4) Credits	notes about the author of the initial version of this file, which was further adapted by the OPTAIN WP3 team

Minimal guidelines

In "**Form**" and "**Form_R**" all elements defined by INSPIRE for "discovery" are listed.
(M) *(C)* — Elements with *(M)* are mandatory, while elements with *(C)* are conditionals
The first column (A) is used to **show/hide** the rows of the form.
yes — Rows where the value is "**yes**" are shown and may be filled.
no — Where the value is "**no**", the correspondent rows are hidden: this is useful for constant values like "contacts" or "language", where the text can be written once and then hidden.

Some cells can be edited as free text: in this case it is sufficient to replace the string with the appropriate text.
Buildings — In other cases, the cells shall contain values taken from dropdown list.

GEMET - INSPIRE themes, version 1.0 — Cells with dark background have been filled in with default values; you may leave them as they are (since requested by INSPIRE)

Tips and tricks — It is useful to start the editing of metadata with "constant" values, like the ones in **cyan**:
* organisation name
* email
* web site
* bounding box (west, east, south, north)
For all these cells, once filled with appropriate values, it is sufficient to write "no" in the first column (A) and then click the "**Show\Hide**" functionality.

In "Form_R" sheet this information does not have to be added, because R adds it automatically.

Some spatial information can be read from .shp and .tif spatial datasets — In "**Form_R**" sheet these cells do not have to be given manually if the file format of the dataset is **.shp** or **.tif**, because these can be read from the GIS files. In the case of any other file formats, e.g.: .txt, .csv, .adf, etc. this has to be given manually.

Figure A3.1. Snippet of the Readme sheet of the Metadata_TG_v1.0.xlsm tool

II. *Form*

The main sheet for filling up the information of a single dataset metadata. It has mostly self-explanatory instructions on what information is needed (**Column C**) with additional comments in cells with a red corner indicator (**Column D**), which becomes visible when you hover your cursor over the cell.

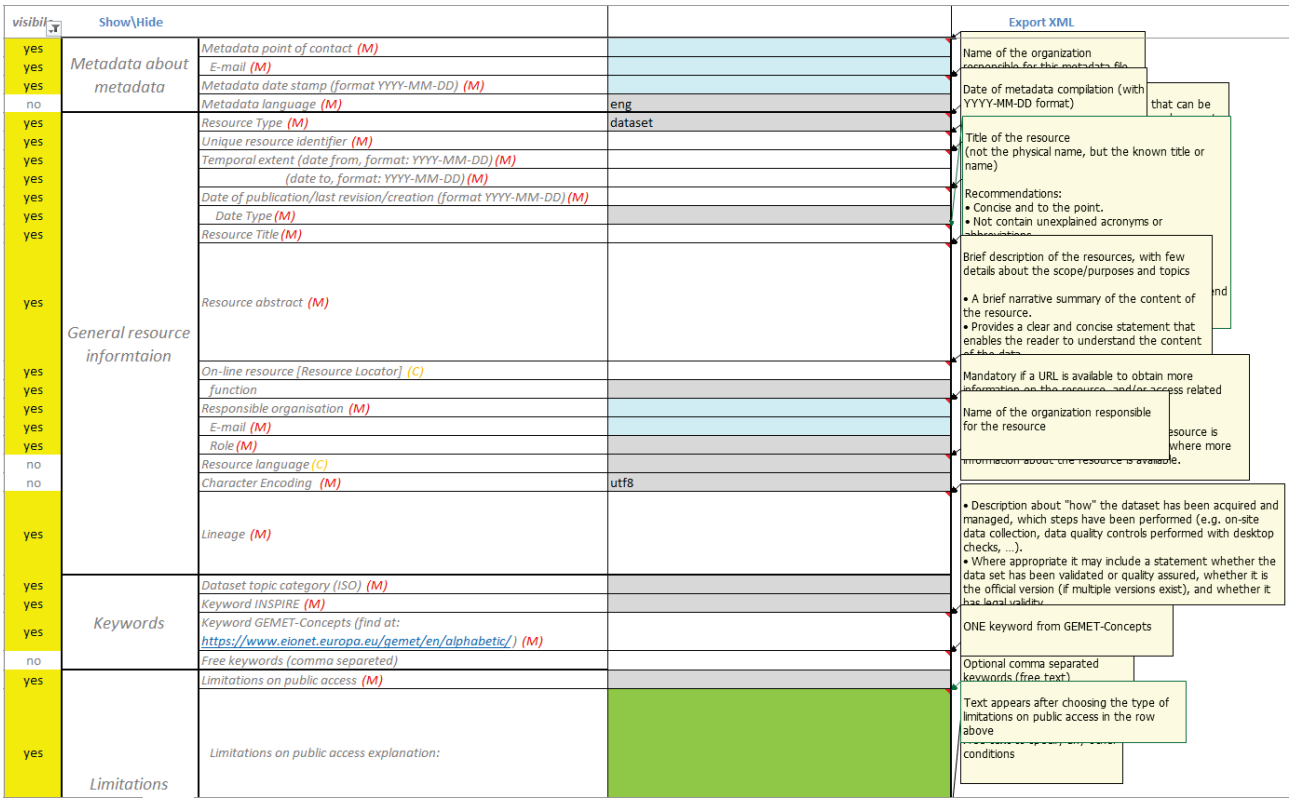


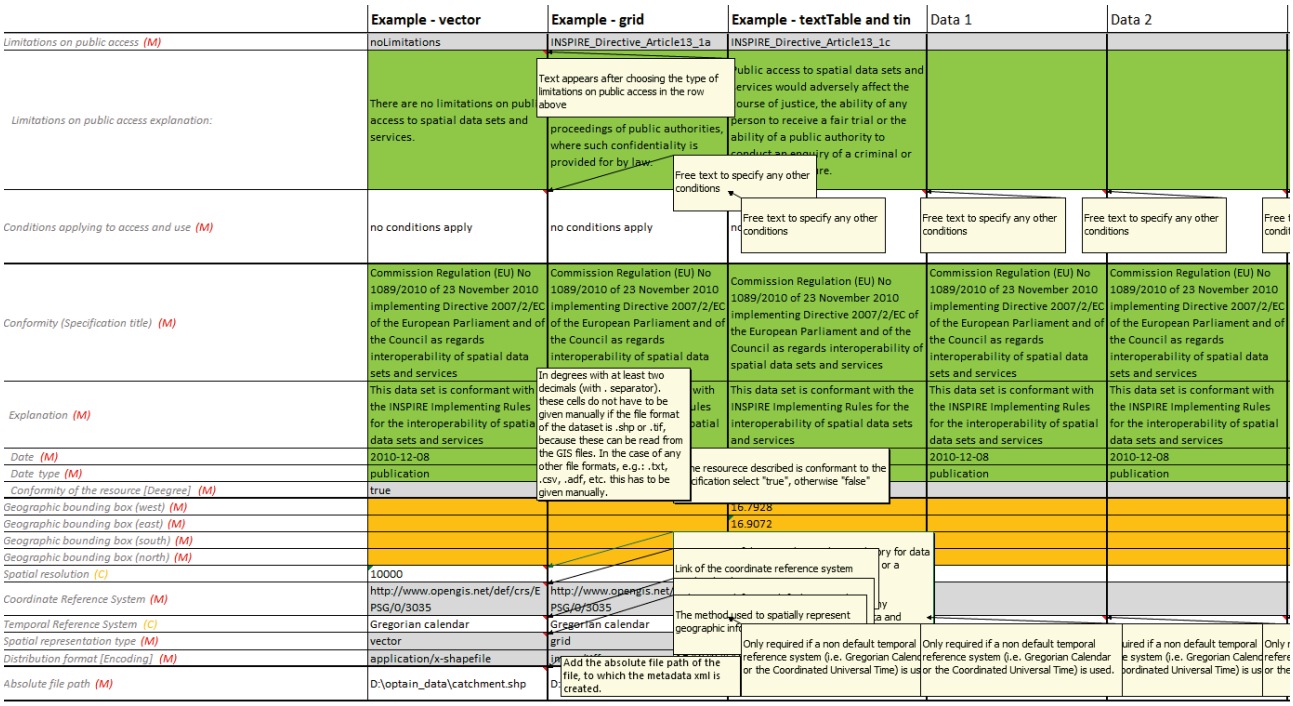
Figure A3.2. Snippet of the Form sheet of the Metadata_TG_v1.0.xlsm tool

The finished metadata information of a single dataset can be exported to an .xml file by pressing the "Export XML" cell and specifying the location for saving the output.

III. *Form_R sheet*

Using this sheet and the provided R script file (metadata_creator.R) it is possible to generate metadata for multiple datasets.

This sheet contains almost the same information as in the Form sheet, although the metadata information of multiple datasets can be provided in separate columns (next to the examples in columns D-F).



	Example - vector	Example - grid	Example - textTable and tin	Data 1	Data 2
Limitations on public access (M)	noLimitations	INSPIRE_Directive_Article13_1a	INSPIRE_Directive_Article13_1c		
Limitations on public access explanation:	There are no limitations on publ access to spatial data sets and services.	proceedings of public authorities, where such confidentiality is provided for by law.	ublic access to spatial data sets and ervices would adversely affect the ourse of justice, the ability of any erson to receive a fair trial or the ability of a public authority to re.		
Conditions applying to access and use (M)	no conditions apply	no conditions apply			
Conformity (Specification title) (M)	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	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	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	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	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
Explanation (M)	This data set is conformant with the INSPIRE Implementing Rules for the interoperability of spatia data sets and services		This data set is conformant with the INSPIRE Implementing Rules for the interoperability of spatial data sets and services	This data set is conformant with the INSPIRE Implementing Rules for the interoperability of spatial data sets and services	This data set is conformant with the INSPIRE Implementing Rules for the interoperability of spatial data sets and services
Date (M)	2010-12-08			2010-12-08	2010-12-08
Date type (M)	publication			publication	publication
Conformity of the resource [Deegree] (M)	true				
Geographic bounding box (west) (M)			16.7928		
Geographic bounding box (east) (M)			16.9072		
Geographic bounding box (south) (M)					
Geographic bounding box (north) (M)					
Spatial resolution (C)	10000				
Coordinate Reference System (M)	http://www.opengis.net/def/crs/EPSG/0/3035	http://www.opengis.net/ PSG/0/3035			
Temporal Reference System (C)	Gregorian calendar	Gregorian calendar			
Spatial representation type (M)	vector	grid			
Distribution format [Encoding] (M)	application/x-shapefile				
Absolute file path (M)	D:\optain_data\catchment.shp	D			

Figure A3.3. Snippet of the Form_R sheet of the Metadata_TG_v1.0.xlsm tool

Complete examples for vector, raster/grid and other file types (e.g. text tables in .csv or .txt and tin in .adf formats) are provided in columns D, E and F respectively.

There are formatted columns presets available (as examples). If the user wants to add additional columns, he/she must copy the preset format of the existing columns. Green cells are defined as the default values and should not be edited.

The **Form_R** sheet has to be filled-in similarly to the **Form** sheet for all the files which metadata XML will be generated.

Please note that:

1. *Metadata date stamp (format YYYY-MM-DD) (M)* is created automatically.
2. *Geographic bounding box (M)* - west, east, south, north - is read automatically from the GIS file if data format is **.shp** or **.tif**. In the case of any other file formats, e.g.: .txt, .csv, .adf, etc. this has to be given manually.
3. *Spatial resolution (C)* is not needed in case of raster/grid files, it is automatically determined. In case of vector or text table files, it is mandatory to give manually an equivalent scale or resolution.
4. *Absolute file path (M)* is mandatory and has to be given correctly or R script will not work.

After the XML file export the user can validate the file according to INSPIRE requirements at: https://inspire.ec.europa.eu/validator/about/ . The "Conformance Class 2: INSPIRE data sets and data set series interoperability metadata" option in Metadata (TG version 2.0) section should be chosen.