Guidelines for new applications

Revision: 2.0

Author(s)/Organisation(s):

Carmelo Attardo, Roderic Molina, Giorgio Saio (GISIG)

Content

Index of Figures

Acronyms

ARe3NA	A Reusable INSPIRE Reference Platform
COTS	Commercial Off-the-Shelf
DB	Database
DoW	Document of Work
EIF	European Interoperable Framework
EU	European Union
Geo-ICT	Geospatial Information and Communication technologies
GIS	Geographical Information System
GML	Geography Markup Language
HTTP	Hypertext transfer protocol
ICT	Information and Communication Technology
INSPIRE	Infrastructure for Spatial Information in the European Community
IPR	Intellectual Property Rights
ISA	Interoperability Solutions for European Public Administrations
OGC	Open Geospatial Consortium
OSS	Open Source Software
OWS	OGC web Services
PIDs	Persistent identifiers
PSI	Public Sector Information
RDF	Resources Description Framework
SEIS	Shared Environmental Information System
TF	Thesaurus Framework
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WPS	Web Processing Service

1 Introduction

This document consists on a guide for the development of further applications within eENVplus.

The document focuses on the methodology applied to design and deploy the eENVplus project pilots and the provision of a set of recommendations for a correct and effective implementation of new applications.

The content of this document is not a user manual of the eENVplus components. The explanation on how to use the eENVplus components is exposed by the Training Framework[1] accessible by the eENVplus web site[2]. This document is mainly a guideline to analyse, design and deploy a new application based on the INSPIRE paradigm and exploiting the eENVplus infrastructure with a collection of recommendations which are not included either in INSPIRE implementing rules and in the eENVplus component training materials.

For example, about the production of metadata, the INSPIRE implementing rules provide the rules to produce a formally compliant metadata. At the same time, the eENVplus tools permit to easily manage the production of metadata and to increase the semantic interoperability through the exploitation of the Thesaurus Framework.

2 From the vision to the implementation

2.1 eENVplus vision

The strategy of eENVplus is based on the demonstration of a real capacity to create added value for its partners and content providers, as well as to offer benefits to stakeholders in the use and exploitation of environmental data exposed according to the EU standards as driven by the INSPIRE directive.

The common perception of the implementation of INSPIRE is the complexity of the models and the protocols, the uncertainty of the compliance of the operational implementation, the unavailability of a set of well-defined procedures able to focus the users on their own needed avoiding them the technicality of the operational implementation of the INSPIRE protocols.

The main goal of eENVplus project has been to hide the complexity of the operational implementation through the customisation and the integration of the available tools, highlighting the users on the real problems which they have to address making the INSPIRE infrastructure really useful and advantageous.

A distributed infrastructure is based on the concept of maintaining the information on the source place, making them accessible by the community in a standardized and shared way in order to have always at disposal the most updated information, accessible and exposed in a way based on shared standard. In the same way, not only the information but also the processing procedures could be maintained and shared between the involved thematic stakeholders using shared operational workflows and based on shared standards processing protocols.

A distributed infrastructure, at any level of extension (local, regional, national, European,….), is composed by different nodes each of them exposing specific resources, from data (mainly data providers nodes) to services and clients applications.

1 http://www.eenvplus.eu/project/eenvplus-training/
2 http://www.eenvplus.eu

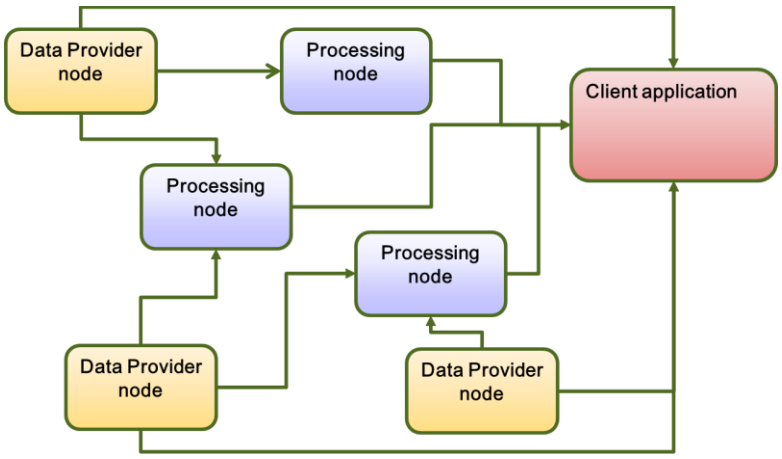


Figure 1 - Distributed infrastructure

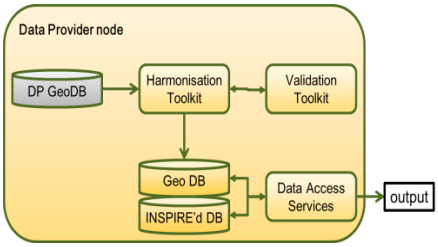


Figure 2 – Sample of data provider node

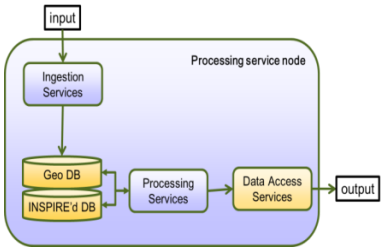


Figure 3 – Sample of processing node

The current operational infrastructure has the peculiarity to be built on protocols, which can be based on different standards, not necessarily interoperable. These can be based on:

- own proprietary standards,
- protocols driven by the used technology,
- de-facto standards protocols,
- open standards

or a mix of these typologies.

The introduction of the INSPIRE Directive and its implementing rules drives the evolution of the spatial data infrastructures (mainly the SDI related to the management of environmental data, due to the fact that the INSPIRE directive refers mainly to the environmental data) to converge to a common protocol or a set of common protocols for each different SDI component and protocol which is defined by this directive. In addition, this protocol is not just focused on network rules, but it also defines the rules for data and metadata structures through the provision of shared common data and metadata models.

Finally, the INSPIRE Directive defines the rules and the protocols at the following levels:

- metadata
- data
- services

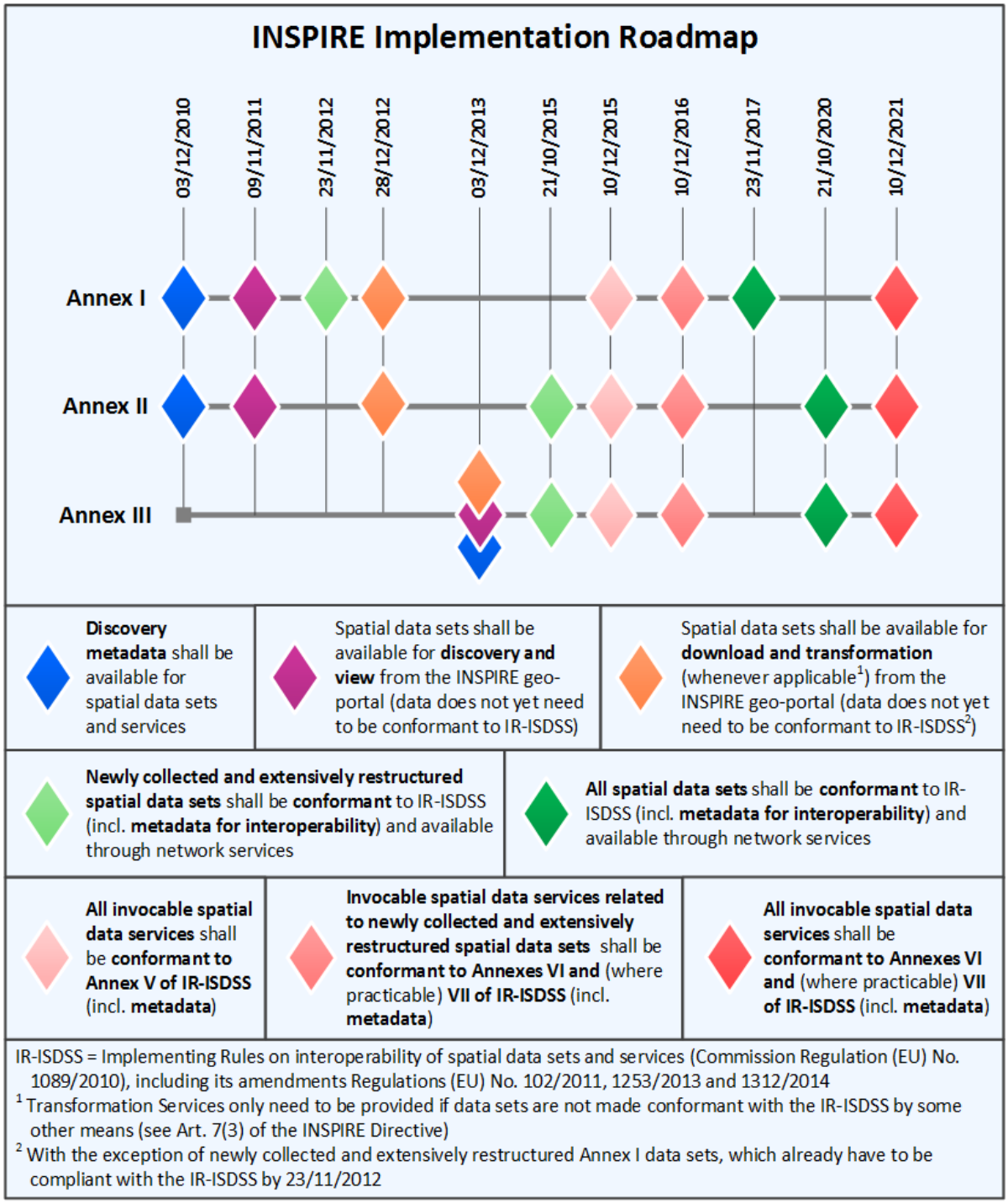


Figure 4 – INSPIRE implementation roadmap

The INSPIRE implementation roadmap, detailed in the above figure[3], details the timeline of the directive implementation which will be fully operational on December 2021. Until now (end of 2015), the Member States concentrated its efforts on the respecting of the first deadlines, mainly related the metadata and a set of services (discovery, view and download services). In this last years, the effort is moving on the data harmonisation, on the understanding how these new models can be integrated into the existing operational processes and how the availability of data exposed with common shared protocols and models can be exploited by the design of new advanced services.

3 Source: official INSPIRE web site http://inspire.ec.europa.eu/index.cfm/pageid/44

eENVplus project focused its efforts mainly in supporting the stakeholders involved in this process (mainly the Environmental agencies) in the operational implementation of the INSPIRE Directive through the provision of:

- a methodological approach in designing/revising operational workflow by the integration of INSPIRE conformant data and services
- a set of operational components able to dress the existing data and processes in order to be compliant to the INSPIRE implementing rules
- a knowledge base supported by a knowledge framework able to learn and to drive the stakeholders into this change management process.

The main result produced by the eENVplus project has been:

- the increased awareness of the stakeholders in “what it is” and “what it means” the operational implementation of INSPIRE rules and protocols,
- how the apparent complexity enclosed in the INSPIRE implementing rules can be overcome by the integration of technical components and by a constant and competent technical support able to transfer the knowledge and to build the capacity to the stakeholders
- the awareness of the stakeholders about what is addressed by INSPIRE and what is NOT addressed by INSPIRE but that have to be taken into account to achieve an effective interoperability of data and services,

This guideline will provide a list of recommendations either for a correct implementation of the INSPIRE implementing rules either for increasing the interoperability of data and services which is not covered by the only application of these rules.

2.2 Implementation: the eENVplus bricks

The advantage of a shared infrastructure exposing interoperable data has been well described and analysed in these last 10 years in many publications, projects, initiatives. Making this kind of infrastructure running and operational impacts at different levels: policy, organisation, technology, knowledge, standards. eENVplus project pursued to give a possible answer addressing all these aspects and simplifying their implementation through the definition of functional processes built on the base of core concepts and technical components able to hide the complexity to the user.

Building and exploiting SDI based on INSPIRE paradigm able to share interoperable data through interoperable services has been realized and differently implemented into the 10 pilot cases of the projects through the combination of the eENVplus components (bricks) representing some of the main core tasks needed to provide such infrastructure.

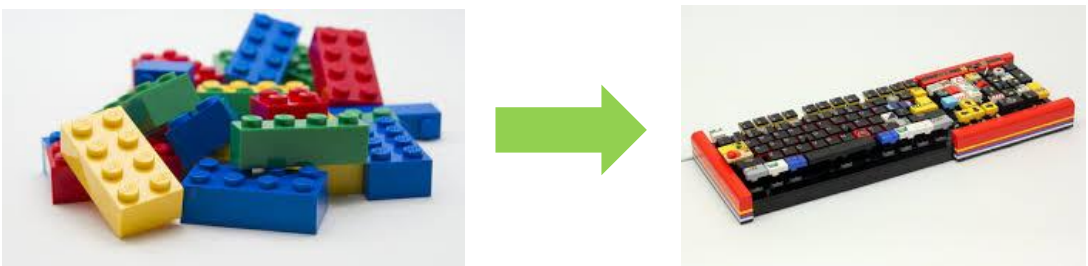


Figure 5 – Using interoperable bricks for building applications

The core tasks addressed by the eENVplus project are displayed into the following picture:

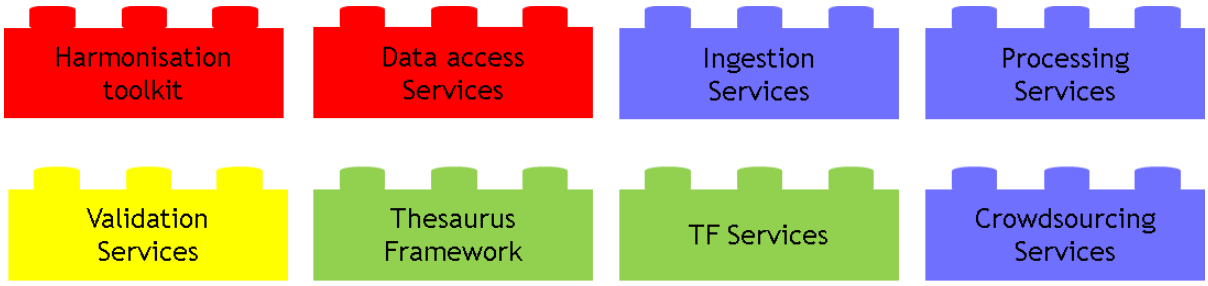


Figure 6 – eENVplus core bricks

The combination of these bricks and the integration of them into the operational environment of the involved stakeholders can facilitate the development of those processes and services, making the new applications (as well as the redesigned data) more interoperable and re-usable.

The proposed bricks don't represent the only possible solution to realize these kind of processes; they represent a possible solution supporting the design and the deployment of applications based on interoperable components and exploiting spatial data infrastructure components based on INSPIRE implementing rules. What is important is the core vision which is on the base of the designed solution, which is mainly based on splitting the data and processes in independent component which can be easily replicated and reused in different contexts, making each of them as a building brick reusable and exploitable by other applications.

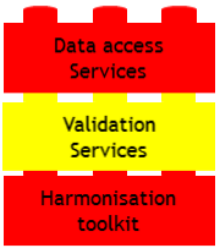
The main characteristic of each brick of the eENVplus Infrastructure is to have the input/output interface based as much as possible on the INSPIRE protocols and rules. This characteristic makes each single brick interoperable with the other bricks and exploitable by the enabled users.

Recommendation 1

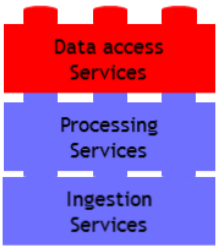
Each component of an application has to be deployed with an input/output interface based on INSPIRE implementing rules

This recommendation has to be applied to data and services. This means that the input/output data have to refer as much as possible to the INSPIRE data model as well as the service exposing or consuming this data have to be based on the INSPIRE standard services. Only in this way each developed component can be connected and integrated with the other components.

The following figure shows two examples of how the interoperable bricks can be combined to provide different kind of services.



Example of combination of bricks for data provider node



Example of combination of bricks for service provider node

Figure 7 – Example of combination of bricks for data providers and service providers

The first one (left side) represents a node of a data provider which needs to:

- harmonise proper data (Harmonisation toolkit),
- validate the harmonised data (Validation service)
- and expose the harmonised data with standard service (Data access service).

The second one (right side) represents a possible implementation of a node of a service provider which needs to:

- consume data service in input to ingest data (Ingestion service),
- process the ingested data (processing service)
- and expose the resulting data with standard service (Data access service).

The combination of the different components into an infrastructure is possible if the protocol to connect them refers to the same standards.

3 Design a new application

Hereafter is described the workflow adopted in the project to understand the single pilot, identifying the different use cases, the functional and not-functional requirements, the design the overall workflow implementing the use cases and the technical specification of each workflow component.

A specific application can be built in a lot of way and there can be different solutions which will provide the requested services. These solutions usually address specific requirements in term of performance, availability, interoperability. Building a new application based on a shared interoperable environment means mainly to drive the solution through the exploitation of the eENVplus bricks, the interoperable data and services available in the infrastructure and the design of the new components with interoperability characteristics.

This approach has been applied and refined during the project to design each of the 10 eENVplus pilots' applications, by designing solutions to different problems in different countries. This heterogeneity of the addressed scenarios (many issues, many countries) highlights the reusability of the described approach also for the design of new applications.

3.1 Use case and requirements identifications

The proposed approach to design a new application starts by its definition through the identification and the detailed description of the different use cases covering the overall application.

After the identification of the use cases, the next step is the analysis of the use cases and the identification of the user requirements. An important and difficult step of designing a software product is determining what the user actually wants.. Often the user is not able to accurately communicate their needs from a technical viewpoint, and the provided information may also be incomplete or even inaccurate. The responsibility of completely understanding what the user wants then falls on the technology providers.

Recommendation 2

The identification of the user requirements and the detailed description of the related functional and not-functional requirements needs a strong collaboration between users (thematic competence) and the technology provider (technical competence) with good knowledge of the eENVplus bricks.

This activity provides the list of functional and not-functional requirements, which will drive the design of the overall operational workflow as a combination of different interoperable bricks.

In the following a brief definition of how eENVplus defines functional and not-functional requirements:

Functional requirements, specifying functions that the system components must be able to perform

They are strictly linked to the eENVplus services:

- Harmonisation Toolkit
- Validation Toolkit
- eENVplus Catalogue and Connection
- Thesaurus Framework (TF) for Metadata Compilation
- TF for Data Discovery
- Web Map Service (WMS)
- Web Feature Service (WFS)
- Web Coverage Service (WCS)
- Catalogue Service for Web (CSW)
- Sensor Observation Service (SOS)
- Web Processing Service (WPS)
- Orchestration Service
- Reporting Service
- Crowdsourcing Service
- Validation Service
- Ingestion Service and Work Flow
- 3D Visualisation Support Service
- Mobile App Support Service

Non-Functional Requirements including:

- Performance requirements, indicating numerical values for measurable variables used to assess applications.
- Operational, User Interface and usability requirements, describing how the system will run and how it will enhance usability issues.
- Reliability requirements, denoting the acceptable mean time between software failures averaged over a significant time period.
- Security and Legal requirements, dealing with the management of users a various levels of access to data.

In the eENVplus project, the collection of the functional and not functional requirements has been performed and structured with the exploitation of a platform permitting the documentation of each identified requirement and tracking the discussion between technical and users in order to.

- better identify and describe the requirement through the maintenance of the discussion tracking
- have a common unique repository for the collection of the requirements which have to be maintained also during the design and the development of the new application. This for two main reason:
    - technical documentation of the designed functionalities
    - knowledge base for driving the test plan of the designed solutions which have to realize the collected requirements

Recommendation 3

Define a well-structured process to collect and to document the functional and not functional requirements, taking into account the reference interoperable services and all those needs and constraints (functional and not functional) which have to be respected by the application. This will drive the design of the application as well as the testing phase.

Following in Figure 8 a snapshot of the tool based on Redmine used to collect the functional and not-functional requirements for the eENVplus project pilots.

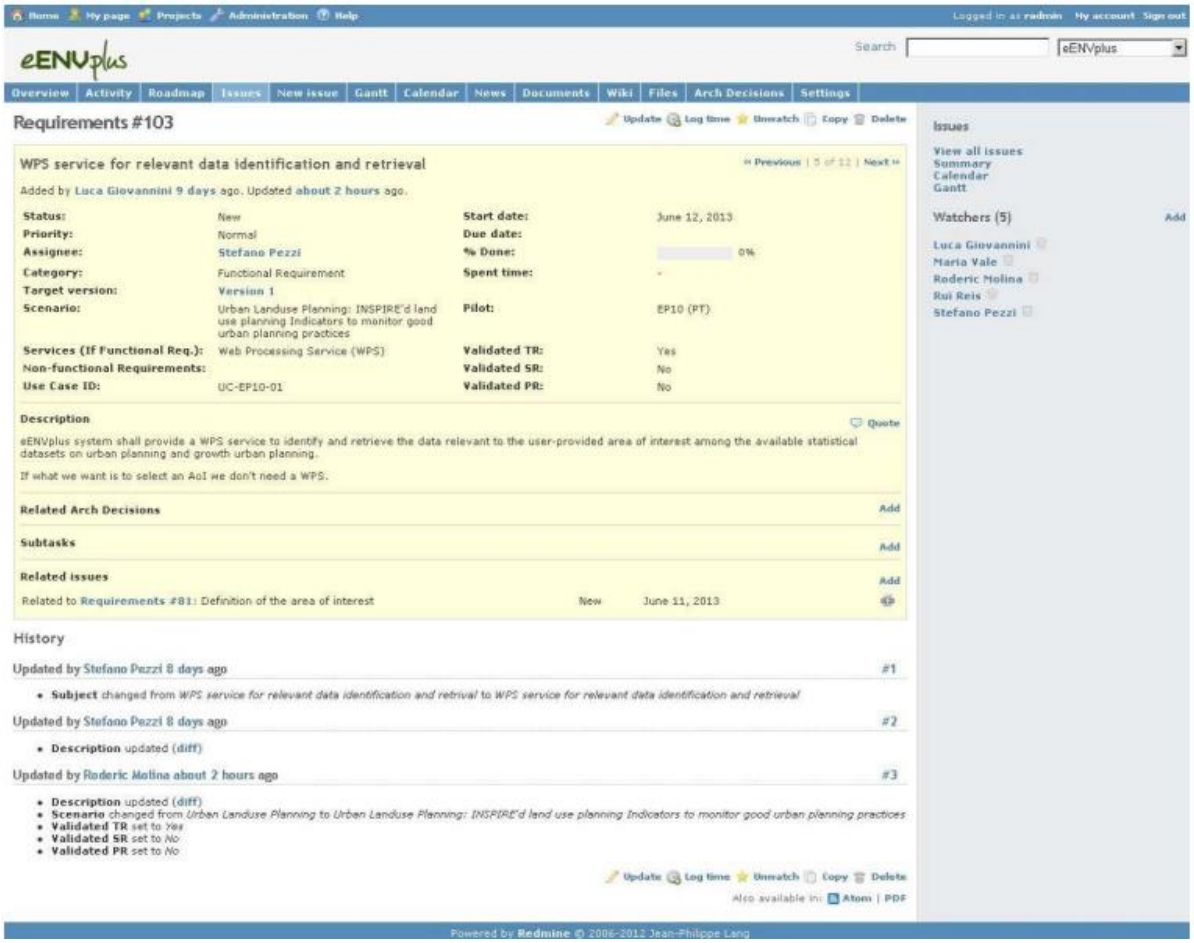


Figure 8 – Tool for collecting eENVplus requirements based on Redmine

3.2 Design the solution and operational workflow

After the collection of the complete list of application requirements, the next step is the design of the solution which have to be based on the interoperability paradigm, taking care of INSPIRE implementing rules and technical guidelines and the availability of the eENVplus bricks which can be exploited and refined according to the identified requirements.

The design of the solutions can be split in three main steps as displayed in the following picture:

Figure 9 – Design the solution

- (meta)data design The identification and the design of the data and metadata structures and the provision of the harmonised data and metadata according to standard services. This deals mainly with metadata and data harmonisation-validation processes, and integration with the eENVplus Thesaurus Framework tools

- Services design The identification of the needed services and their customisation or new implementations. This have to take into account either the functional but, overall, the not-functional requirements, overall in term of availability and efficiency

- Client design The design of the application interface (client development) able to implement the operational workflow

In order to better understand the interaction between data, services, client interfaces and to have an overall schema representing the defined solution, each of the identified components (data, services, client interfaces) will be interlinked into an operational workflows able to schematize the interaction between them. The proposed approach in building this workflow is to include into this schema the involved users, the involved data (specifying who exposes these involved data, i.e. data providers), the involved services (specifying which expose these involved services, i.e. service providers).

Recommendation 4

Identify and describe all the possible interactions between the different components implementing the application (users, data and services), schematizing them in a workflow schema in order to have an overall vision of the designed application.

An example of such workflows is reported in the following picture (eENplus pilot “A SEIS for air quality data”):

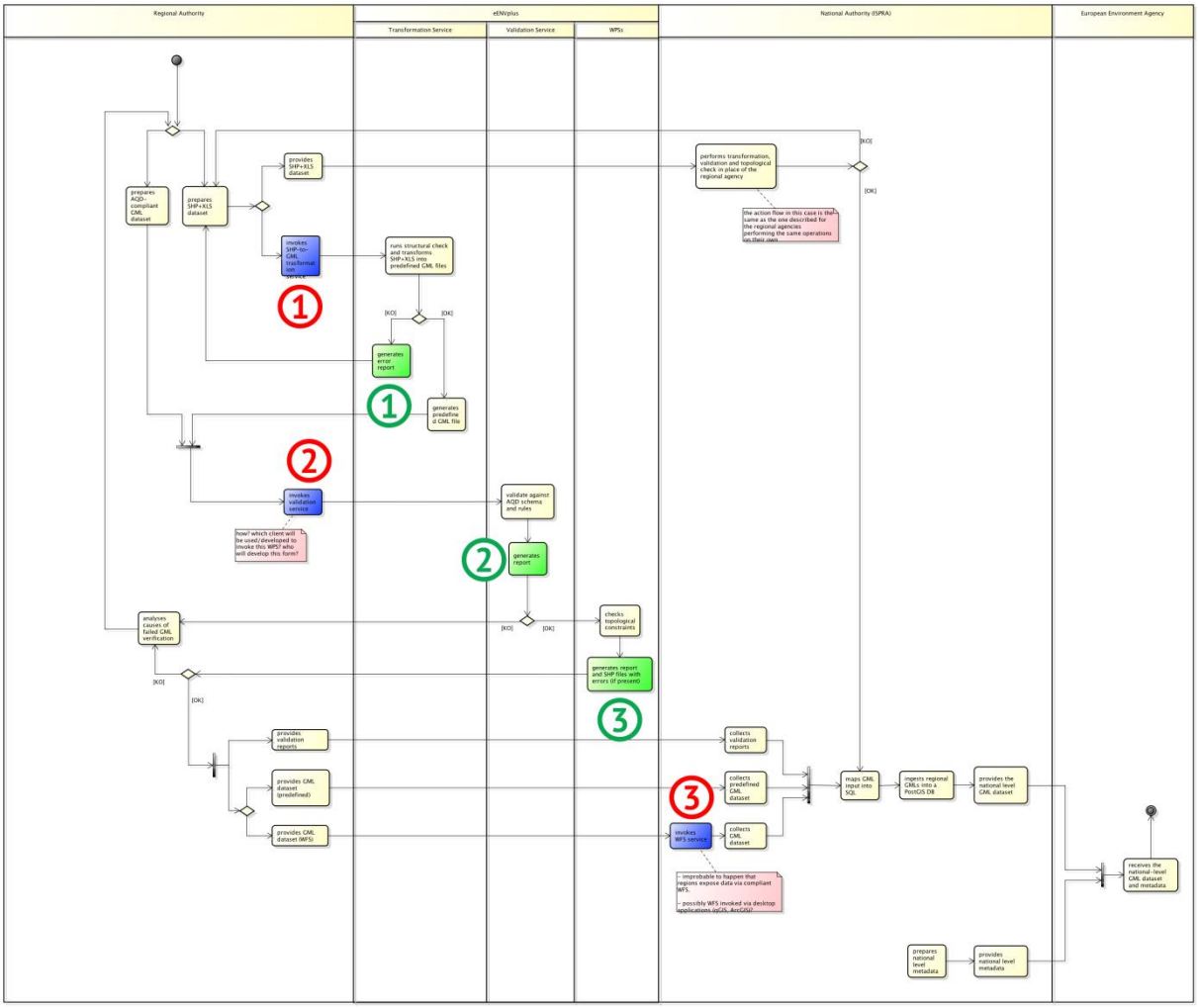


Figure 10 – Pilot workflow

The above workflow is supported by a set of detailed informations which define for each use case identified into the application (identified into the workflow schema by the coloured numbers):

- input and output data
- a detailed description of each step enclosed into the use case workflow
- the mock-up of the interface (if any) exposing the user interface
- the list of services exposed by the infrastructure (eENVplus bricks or other existing available components) which will be integrated (with eventual customisation) to realise the process
- the list of services not exposed by the existing infrastructure which have to be developed to complete the processes.

Examples of this specification (workflow, input/output data, services, mock-up) related to the developed eENVplus pilot applications are reported in deliverable “D7.2 Applications system specification”.

4 Metadata and Data definition

4.1 Understand the data needed to the process

As specified in the previous chapter, one of the main activities is the identification of the data (and related metadata) needed to the new application. This action is driven by the context in which the new applications have to be designed, i.e. an operational distributed infrastructure where data and services are exposed according to standards models and interfaces. For this reason, in order to respect this assumption, it's needed:

- to identify the data needed for the application
- for each of this information, to identify the shared data models (mainly referred to the INSPIRE data themes) in which each of this data is enclosed
- if it's not enclosed in any of the INSPIRE data model, to plan an extension of one of the involved data model which can better include the information

This action have to be performed both for the input data and for the output data.

It is also needed to perform this alignment between the requested data for the new application and the INSPIRE data models,

Recommendation 5

it's recommended to monitor the INSPIRE website in order to ensure the alignment with the most updated version of the INSPIRE implementing rules and Technical Guidelines as well as to check the on-going discussion and open issues published in the thematic cluster related to the involved data themes.

4.1.1 Find reference data models for the scope

One of the way to identify the correct data model relevant for the new application is to exploit the tool "Find your scope" published on INSPIRE official website in the section "INSPIRE Interactive Data Specification"[4].

The application "Find your scope" supports data providers with identification of the INSPIRE spatial data themes and spatial object types that are relevant to the dataset(s) they manage. This application is foreseen to be useful especially in situations when datasets fall under two or more INSPIRE data themes / application schemas content.

The main outcome of this application is the list of the INSPIRE objects, including their properties – attributes, code lists values etc. which are relevant to the application. The final list also includes all associated objects and their properties.

In addition, a user after completing the whole interactive workflow will be also able to:

1. Identify data properties needed to be transformed according to INSPIRE definitions.
2. Identify missing information in his/her dataset that is required by INSPIRE object definitions. (Gap Analysis)
3. Identify a potential INSPIRE data theme extension, to cover his/her full dataset scope.

Recommendation 6

The use of “Find your scope” tool published on INSPIRE official website guarantees the alignment of the last INSPIRE data models and it’s a valid support in understanding the INSPIRE data models and discovery the corrected data components to reach the INSPIRE compliance

4.1.2 Data model extension

The identification of the correct data objects able to include the complete set of information needed for the application has to be deeply analysed and understood in order to define if the selected structures fit the requirements needs.

Recommendation 7

A deep analysis of the technical specification of the selected data models has to be carefully performed to better understand the data structures and the correct understanding of the described implementation to correctly map own data into the selected models.

Recommendation 8

A good support in understanding the selected data models is provided by discussions enclosed into the INSPIRE thematic clusters. These communities are a concrete operational support to perform a correct mapping of the data.

The identification of the correct data objects able to include the complete set of information needed for the application often implies the definition of some extension to the selected models which have to be defined and applied in order to satisfy the application requirements.

It’s possible to identify two main classes of data model extensions:

- the ones which include new elements into the model (new feature classes, attribute, fields)
- the ones which modify rules and constraints applied to the existing data model elements.

The first type of extension have a strong impact in the process because this strongly impact on the harmonisation process and, in detail, in the schema defining the data model (XSD files used to harmonise the data), the structure of the matching table, the validation service customisation.

The second type of extension, which impact exclusively to refinement of some constraints without modifying the structure of the data model, has surely a less impact on the overall harmonisation process. This is the case of changing some fields to be mandatory or the redefinition of some fields domains (for example, value which have to be into more restricted fields of value or be selected from additional code lists).

In this case, there could be two different approaches to manage this kind of extension:

- changing the XSD schema including this additional constraints
- customize the validation service/process without modifying the XSD schema.

The adoption of this second approach doesn’t impact on the data models definition files but just only on the validation process.

4 http://inspire-regadmin.jrc.ec.europa.eu/dataspecification/FindYourScope.action

Recommendation 9

Limit as much as possible the extension of models, overall those extensions which impact on the data structure. Check carefully the mapping into the INSPIRE data themes of the needed data.

After the accurate analysis of the available data structures and the impossibility to avoid structural extension, before design the structural extension of the selected data objects, it's suggested to refer to the related thematic community (INSPIRE thematic clusters) to check if the planned extensions have been already analysed by the thematic community and if it have already provided suggestions and solution to apply these extensions.

In any case, it's suggested to submit to comments/suggestion the proposed extension to the thematic communities, refine them according to the collected feedbacks and share the final solution which will be used.

Recommendation 10

In case of data model extension:

- check the reference thematic community (INSPIRE thematic clusters) if similar extension have been already proposed/discussed/solved

- communicate the INSPIRE thematic cluster community about additional extensions to share them with the thematic communities.

4.1.3 Data harmonisation

The referenced methodology proposed and applied by the exploitation of the Harmonisation toolkit delivered by the project has been designed and refined for the harmonisation of the overall set of data involved into the eENVplus project pilots. The experience made in the project highlight the need to spend the needed time for a complete understanding the target schema, for the detailed compilation of the matching tables before implementing the effective harmonisation process using the selected tools.

A common mistake in data harmonisation is starting from scratch to use the programs performing the data transformation without before defining the mapping rules.

Recommendation 11

Designing the data mapping using the matching table and filling them as much detailed as possible provides a better understanding of the overall harmonisation and saves time in the last step of the mapping implementation through the selected software tool.

When a new process of data harmonisation have to be performed, it's important to take into account the following recommendation:

Recommendation 12

In the harmonisation process of a source schema versus an INSPUIRE target schema, it's important to map all the information from the source schema to the target schema, independently by the fact that

these information are classified mandatory or not in the target schema. In other words, don't limit the filling of the mandatory fields but provide all the available information.

This aspect it's important for two main reasons:

- An harmonisation process performed to address a specific request have to be exploit also for other potential processes, first of all, the obligation driven by the INSPIRE roadmap.
- If the same datasets are requested by other processes, it doesn't need to rebuild the harmonisation but just to refine them in order to satisfy the overall requests.

For this reason, it's important to fill all the available information into the target schema interpreting the attribute "mandatory" not as the only information to be provided, but as the minimum information to be provided.

Particularly relevant has been the experience gained by most of the project partners to implement data harmonization processes, using powerful tools such as HALE[5] and effective procedures based on detailed matching tables. This experience constitutes a valuable intangible asset, that can be re-used in several scenarios, ranging from the optimization of specific INSPIRE related business processes carried out within the partner organizations, to the creation of new specialized consultancy services to be offered by private partners in the Geo-ICT market.

4.2 Metadata management

The provision and the management of the metadata into a spatial data infrastructure have to take care of different aspects which can affect a lot the final expected results of this topic. The main functionalities requested to a metadata catalogue is mainly the capacity to discover the data collected into the catalogue and to provide to the users all the information for its use.

This capacity is not completely covered by the compliance to INSPIRE: building an INSPIRE compliant catalogue with INSPIRE compliant metadata does not guarantee the operational capacity to discover the requested data and services.

The INSPIRE compliance drives the data provider to configure and deploy an interoperable catalogue service exposing standard metadata but this often is not sufficient to reach the requested needs.

There are some aspects which have to be taken into account in design and provide interoperable metadata which are:

- The completeness and the correctness of the information which have to be guaranteed by well-defined rules for filling some free-text metadata fields (like abstract, lineage,…)
- The selection of the keywords to describe data and to search for data which should mainly be:
    - Language independent
    - Shared by the different thematic communities interested to the described data.

This last aspect refers mainly to the semantic aspects of the content defined into the metadata.

Recommendation 13

The compilation of the metadata fields have to be detailed as much as possible, overall taking into account the possible applications where this data can contribute. Even if a lot of metadata fields are

5 The HUMBOLDT Alignment Editor (http://www.dhpanel.eu/humboldt-framework/hale.html)

defined as not mandatory by the reference regulation (INSIPRE/National/Regional), it's suggested to fill as much information as possible, interpreting all metadata fields as mandatory.

Another key aspect in compiling metadata are the keywords. It's important to define as much number of keywords as possible, listing a set of keywords answering to the following questions:

- INSPIRE data themes
- General contexts (for example: forest, land, water, urban planning….)
- Reference programmes and actions for which the data are provided (for example: Water Framework Directive, Habitat directive, National/Regional programmes, etc…)

In addition, in order to provide a uniform description of data, it will better to not use free text keywords but recovering the keywords from existing code lists and thesauri. This can be facilitate through the use of the Thesaurus Framework LusTRE developed in eENVplus: the knowledge infrastructure of linked thesauri with the set of services for its exploitation within client applications, enabling the discovery and the selection of the keywords from shared thesauri.

Recommendation 14

Describe data in a uniform way by selecting keywords from existing shared codelist and thesauri, avoiding as much as possible the use of free text keywords.

This aspect increases the interoperability:

- cross-language thesauri, permitting to exploit multi-lingual functionality exploiting the Thesaurus Framework services (LusTRE-ES) integrated into catalogue client
- semantic interoperability: through the exploitation of LusTRE-ES by the catalogue client, it will be possible to discover data searching by keyword, its synonyms, narrower concepts, etc….

About the use of the Thesaurus Framework (its linked thesauri and the services for their exploitation in the context of the management of metadata, please refer to the eENVplus showcase, and in details into the following section:

- Environment Thesaurus: a Linked Thesaurus fRamework for Environment (LusTRE) aims to provide a common terminology solution to face with the multilingual and multicultural issues in environmental data sharing[6].
- Metadata editor: demonstrating the exploitation of the Thesaurus Framework services for the selection of the metadata keywords from the available linked thesauri addressed by LusTRE[7].
- INSPIRE geoportal and the integration with Thesaurus Framework, demonstrating the exploitation of the Thesaurus Framework services into a catalogue client to increase the discoverability of metadata[8]

6 http://showcase.eenvplus.eu/client/thesaurus.htm
7 http://showcase.eenvplus.eu/client/editor.htm
8 http://showcase.eenvplus.eu/client/geoportal.htm

4.3 Data and metadata validation and publication

4.3.1 Data and metadata validation

The harmonisation of data and metadata according to a reference target schema representing an international reference standard like the INSPIRE implementing rules needs a set of procedures to guarantee the compliance of the produced data and metadata to these rules. The validation services provided by the project fulfil this expectation through the exploitation and the extension of official existing services.

Detailed presentations about the validation services provided by the project and used by the eENVplus partners are published on the following sections of the eENVplus showcase:

- Validation service: The eENVplus Validation Service provides executable tests to evaluate whether a GML dataset fulfils the requirements of the INSPIRE Directive[9].
- Metadata editor: demonstrating the integration of the Thesaurus Framework services for the selection of the metadata keywords from the available thesauri addressed by the Thesaurus Framework[10]. The Metadata editor exploit and extend the validation service exposed by the INSPIRE geoportal, providing additional checks related to the metadata for interoperability

In general, there are two main classes of validation procedures. The first one is focused on the validation of data according to reference schemas and rules (for example, the compliance of data and metadata to the INSPIRE implementing rule). In this class these are the validation services described by the Validation service section of the showcase listed above.

The second class of validation procedures are not related to the data structure or the data format but they are strictly related to the validity of the data content. For example, a coverage representing the land cover of a portion of area have to not present overlapping polygons or holes in the map, as well as each polygon have to be labelled with a valid code, etc. These aspects are not related to the format or the structure but they have to be respected at the same way to make the data useful and exploitable.

For this reason, the process for the production of harmonised data have to include validation services to guarantee the correct format and also the correct content of the data to guarantee the data quality to the potential users.

Recommendation 15

Data harmonisation process have to provide the procedure and the related services to guarantee the compliance of the data to the target schema/format as well as the correctness of the data content.

Within the eENVplus project, it has been developed some processing services just to check the data quality in term of content like, for example, the validation service provided for the pilot EP02 on air quality to check the geometry of the provided data[11].

It's clear that the validation procedures to guarantee the data quality are strictly related to the nature of the data. It's also true that building this kind of validation services in the context of an interoperable and shared infrastructure based on reference standards makes this type of services re-usable by the infrastructure users. In this way, the building of specific services able to operate on data exposed through standard service and according to standard models permits to apply them on those dataset

9 http://showcase.eenvplus.eu/client/validation.htm
10 http://showcase.eenvplus.eu/client/editor.htm
11 Pilot "Implementation of a SEIS for Air Quality Data in Italy" - "http://showcase.eenvplus.eu/client/ep02.htm

exposed according to the same standards. Referring to the previous example where the service check the geometry of a land cover datasets, if the data are exposed according to the reference INSPIRE data structure (LC) and the service has been built with the capacity to ingest standards datasets, this service can be applied on each LC INSPIRE dataset.

Recommendation 16

The design of specific validation services to guarantee the quality of data content have to be realized taking care to be in a shared and interoperable infrastructure and, than, providing those interface based on interoperable standards like any other processing services.

4.3.2 Data and metadata publication

The publication of metadata is a well-known process, due to the INSPIRE obligation, active from 2010, establishing that the discovery metadata shall have to be available for data and services. The configuration of catalogue services for distributed infrastructures can be designed and structured organizing them in federated catalogues addressing each request to all federated catalogues endpoints (federation approach) or harvesting in a unique catalogue endpoint the overall set of metadata published by each federated catalogue. In this second approach, it has to be designed an updating procedures which maintain the central catalogue as much updated as possible (this often depends by the updating rules of each federated catalogues).

While the federated approach maintains just only one copy of each metadata without duplication (than, easier to maintain), the harvested approach provide a best performance for each searching request.

A good compromise is the provision of the harvested approach with a strong and well monitored harvesting system, able to maintain all the catalogue nodes updated.

An additional suggestion to drive the user to better find and overall consume the discovered data is the correct and complete documentation of the services exposing the data. It should be provided by the publication services the information about the data but also the related information about the services available to recover the data of interest.

Recommendation 17

Organize and detail in metadata or in the publication service the details about the services exposing the data of interest (WMSs and WFSs services) to easily drive the services consumers (users) to easily discover and exploit the discovered data.

About the publication of data according to the reference standards, the solution suggested by the eENVplus project is based on the data ingestion and publication bricks based on deegree open source software. The main characteristic of this solution is to be integrated with the harmonisation toolkit, enabling the publication of a harmonised GML file through the direct ingestion of it and the quick configuration for the exposition of standards services like WMS and WFS[12].

The recommendation on publishing data through interoperable services is the balancing of the number of services and the amount of data exposed by each of them. Splitting the data into too many services can constrain users to open a lot of connections to recover the interested data and the data providers

12 http://showcase.eenvplus.eu/client/ingestion.htm

to manage and to maintain a lot of resources. On the contrary, to overloading one single service can affect the efficiency and the availability of the service itself.

Recommendation 18

Organise data exposed through WFS balancing the number of the exposed WFSs and the loading of each WFS

In addition, it should be better to define some technical tools avoiding the user to send requests that can block the services making it unavailable for other users, Download services based on WFS should be used to query subsets of data either in term of extent either in term of number of features. The requests which need the complete set of data should be served by atom implementation of download service where the complete set of data is available through standards file transfer service.

Recommendation 19

For dataset with big amount of data, it's recommended to provide together WFS service also the atom implementation based on file transfer standard for the download of the complete set of data

5 Services implementation

Services design

The provision of interoperable service is based on dressing these services with shared and interoperable standards enabling the services to be interconnected between them. The approach proposed and applied by eENVplus project has been to build the processing service taking care overall of the service interface splitting the services in different modules making them interoperable and re-usable. The detailed information about the processing services built in eENVplus are published on the different pilots (following listed) which exploit their capabilities and functionalities:

- A SEIS for Air Quality Data in Belgium[13]: reporting service
- A SEIS for the quality data in Italy[14]: processing services for quality control of coverage
- Forest fire management[15]: processing service for routing calculation
- Geological map harmonisation in Italy and Slovenia[16]: processing services to compute environmental indicators (Landslide susceptibility map)
- Land cover change detection and planning indicators[17]: processing service for coverages intersections

According to this approach, each of the services above listed has been designed as composed by the following modules:

- An input module to ingest the data needed to the process
- The processing module performing the requested operations
- An output module providing the computed output data in a sharable and standard format

13 http://showcase.eenvplus.eu/client/ep01.htm
14 http://showcase.eenvplus.eu/client/ep02.htm
15 http://showcase.eenvplus.eu/client/ep06.htm
16 http://showcase.eenvplus.eu/client/ep09.htm
17 http://showcase.eenvplus.eu/client/ep10.htm

The advantage to have interoperable services is to have the service not linked to a proper data structure and interface but, using open data structure and open interface, to give the possibility to the entire community to quickly replicate the exposed service or to exploit the service directly by the node exposing them (if opened to the community).

Recommendation 20

Design a new service as a combination of three main modules: data ingestion, data processing and data exposition. This approach has to make independent the processing module by the input/output ones making the service independent by the data format.

In addition, to guarantee the interoperability and the re-usability of the services in different contexts,

Recommendation 21

the input as well as the output reference data formats and models of a processing service have to refer to international shared standards, like the INSPIRE data models and related extension.

The capacity of a processing service to be interoperable and, overall, re-usable by different stakeholders and users, implies to take into account some aspects which can make the service more functional and efficient. Processing services consume computational resources of servers and a not-well management of these resources can limit the exploitability of the service itself. The service has to be designed to avoid that big amount of input data or complex and particular data composition could generate long processing time, as well as the block of the service. In a shared environment, this situation is more difficult to preview. A set of constraints in input data and conditions have to be forecasted in order to avoid blocking situation or situation which can generate computational situation with never-ending loops.

Recommendation 22

Design the service avoiding blocking situation and/or computational never-ending loops. Each services should be able to provide an answer (correct or error message) in a time declared into the service metadata.

In this way, each service consumer can know limit and constraints of the exploited service.

For those services which need a lot of time to produce the expected result, it's needed to design the service in asynchronous mode. This means that, when the service is invoked, the server has to be able:

- to keep in charge the input data and the instruction for the requested process,
- to provide to the user a notification confirming the reception of the correct instruction,
- to free the user client in order to permit to the user to perform other operations while the process is running
- to notify to the user the complete execution of the requested service and to provide the instruction to recover the expected output.

These considerations bring to suggest some recommendation in designing and using interoperable services.

Recommendation 23

- dimensioning of the server hosting the service to support the planned overloading
- setting operational constraints to avoid to lock the service: define limit for quantity of data to process, plan escape solution based on computational time consuming
- setting up load balancing and fault tolerance mechanism to guarantee the availability and the operations of the exposed service

In case of asynchronous services, the additional recommendations have to be taken into account into the design and the development phases:

Recommendation 24

Processing chains based on distributed infrastructure and designed in an asynchronous way

➔ need to have a notification service to communicate to the user the progress of own service instance

➔ need to manage concurrent use of these services.

6 Application implementation

Client design

The design of the client application will be based on the and not functional requirements as defined in paragraph 0 "

Design the solution and operational workflow" of this document. The design of simple mock-ups helps a lot the implementation of the client functionalities, and it has to be the main reference for the implementation of the requested functionalities. It's important to take into account that the client interface is the main tool to drive the interaction between data and services and its functionalities have to drive the user to the correct use of them, hiding the complexity of the technical implementation of some reference standards.

Being the system based on interoperable services, it's suggested not to start the development from scratch but to design the interface on already existing framework libraries and/or platforms which already implement a lot of base functionalities. Depending by the technological constraints, the selection of the appropriate technology is driven mainly by the technical environment on which is based the node hosting the client tools (based on OSS or COTS, based on specific technologies and other network constraints).

Recommendation 25

Independently by the infrastructure technology constraints, it's suggested to define the development environment for the client dependently either by technological constraints, either by the requested functionalities, deeply analysing the existing framework/libraries which can help and drive the development of the designed solution.

For example, in the eENVplus project it has been exploited the solution based on Mapstore, an OSS client application with a lot of standard services interfaces already developed and customisable

according to the defined requirements[18]. This tool has been customised to demonstrate an alternative re-use of the overall services instantiated for each pilot application.

The different processes and the related interfaces needed to their use can be constrained by the performance and availability of the different services involved into the processes. Not being these services directly controlled by the service consumer, as previously described (see previous chapter 5 “Services implementation”) for some services it’s necessary to manage workflows based on asynchronous services and the client interface have to be designed to provide the correct tools able to manage these situations.

The eENVplus “support brick” to manage asynchronous services is mainly based on the exploitation of the notification service integrated into the processing service implementation to manage the operational workflow and the communication between the services providers and the service consumers (i.e. the user).

7 Support to the operational implementation: the Training Framework

As preliminary presented at the beginning of this document, this guideline focuses mainly on recommendations and suggestions on the plan, the design and the deploy of new application based on interoperable services exposed into an shared standards based infrastructure, mainly focused on the INSPIRE implementing rules and guidelines.

The set of components (bricks) produced by eENVplus project has been referred into this guideline, without presenting how to setup and how to use them from the implementation point of view. The detailed information about the configuration and the application of the proposed components can be found into the showcase and, overall, in the Training Framework.

The Training Framework is the repository of the base and specific knowledge acquired during the project, able to drive the user in different training paths focusing on the different aspects which have to be addressed to design and deploy an operational interoperable application.

The training modules available in the eENVplus Training Framework has been organised in the following way:

- Background Knowledge, “Knowledge on Directives/ Technologies”. These modules aim at the presentation of the directives, the reference standards and the reference technologies addressed by the project and the project pilots.
- Thematic Knowledge, “Knowledge on specific related knowledge areas”. These modules aim at the explanation of the thematic field of applications addressed by the project pilots, like air quality reporting, forest file management, geological hazards, etc…
- The eENVplus Infrastructure. “Knowledge on the technical outcomes of eENVplus”. These modules aim at the detailed presentation of the components (bricks) designed and developed in the project.
- The eENVplus scenarios. “The implementation of the eENVplus project pilots”. These modules aim at demonstrating the use of the eENVplus structure in different application domains, with transfer of knowledge about the scenario solutions and the use of the eENVplus services within the respective pilots.

18 http://showcase.eenvplus.eu/client/client.htm