FP7-INFRASTRUCTURES-2012-1

Grant Agreement no. 312845

Scoping Study for a pan-European Geological Data Infrastructure

D 4.3

Report on infrastructure needs: Report on infrastructure (hardware, software) and skills training needs for individual NGSO

Deliverable number	*D4.3*
Dissemination level	*Restricted*
Delivery date	*16 May 2014*
Status	*Final*
Author	*Jean-Jacques Serrano, Sylvain Grellet, BRGM*

SEVENTH FRAMEWORK PROGRAMME

EGDI-Scope Project, WP4

Report on infrastructure needs

Table of Contents

List of figures

List of tables

1. Overview of WP4 – Technical design

Work package 4 of the EGDI-scope project sets out the requirements for technical design, deployment and maintenance of a possible European Geological Data Infrastructure (EDGI), in order to fulfill the user requirements and required data provision, identified in WP2 and WP3 (and proceeding parallel to WP4). Some of the most important requirements for an EGDI infrastructure will be based on the principles and directives defined within the INSPIRE framework and other large initiatives dealing with geospatial information. It will be built on the experience of the design, implementation and operations of the different portals and other geological information systems developed within previous and on-going projects and initiatives.

This document is the report on infrastructure needs: Report on infrastructure (hardware, software) and skills training needs for individual NGSO.

The objective is to design an interoperable system supporting a distributed implementation to facilitate the deployment and the operational performance of the system.

The infrastructure needs are based on
- Technical and functional requirements provided by WP2 (D2.4 User needs and functional requirements);
- INSPIRE requirements;
- And elements described in the D4.1 "Report on evaluation of existing interoperable infrastructures" and D4.2 "Report on data models, vocabularies, service and portal specifications for the EGDI".

Each section of the D4.3 provides a link to requirements defined in the D2.4 as described as below:

> *Links to user needs and functional requirements (WP2 - D2.4)*
> *The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:*
> ***Requirements #1, 2, 6***

The overall view on the relationship between architecture needs and D2.4 Requirements is provided in the next table.

Links to user needs and functional requirements identified by WP2 in the D2.4

2. EGDI Architecture viewpoints

To describe the EGDI Architecture we suggest following the ISO RM-ODP (Open Distributed Processing — Reference model. The RM-ODP standards are already used in various geospatial architectures (e.g., the ISO 19100 series of geographic information standards).

Among contributions provided by RM-ODP for the development of open distributed systems are the following:

- It offers a coordinating framework for the standardisation of ODP which is able to integrate current and future standards and maintain consistency among them;

- It offers a conceptual framework and an architecture that integrate aspects related to the distribution, interoperability and portability of systems, in such way that hardware heterogeneity, operating systems, network features, programming languages, databases and management report systems are transparent to the user. Thus, RM-ODP manages the system complexity through a "separation of concerns", i.e. it addresses specific problems from different points of view related to a particular set of concerns.

- It provides a concise specification of concepts defining semantic, methodologies, tools and processes used for the development of ODP.

RM-ODP defines five viewpoints to address the various concerns in developing architectures. A RM-ODP viewpoint is a subdivision of the specifications of a complete system, established to bring together those particular pieces of information relevant to some particular area of concern during the analysis or design of the system. Although separately specified, the viewpoints are not completely independent; key items in each are identified as related to items in the other viewpoints.
However, the viewpoints are sufficiently independent to simplify reasoning about the complete specification. The mutual consistency among the viewpoints is ensured by the architecture defined by RM-ODP, and the use of a common object model provides the glue that binds them all together.

A summary of RM-ODP Viewpoints is provided in Table 1.

Viewpoint Name	Description of RM-ODP Viewpoint
Enterprise	Articulates a "business model" that should be understandable by all stakeholders; focuses on purpose, scope, and policies.
Information	Focuses on the semantics of the information and information processing performed, by describing the structure and content types of supporting data.
Computational / Service	Service-oriented viewpoint that enables distribution through functional decomposition of the system into objects that interact at interfaces.
Engineering	Description of the distribution of processing performed by the system to manage the information and provide the functionality.
Technology	Identification of component instances as physical deployed technology solutions, including network descriptions

Table 1 - RM-ODP viewpoints summary

In the context of EGDI the Computational Viewpoint of RM-ODP has been renamed Service Viewpoint.

For each viewpoint (except the Enterprise Viewpoint) a section is dedicated to the maintenance of the elements addressed in the viewpoint.

3. Enterprise Viewpoint

This viewpoint should describes a "business model" understandable by all stakeholders; focuses on purpose, scope, and policies.

EGDI is THE European counter for serving geological data, models and derived thematic products. It delivers reliable, up-to-date, interoperable geological data and information covering all European countries (or greater sections of it) in the domains of geology, geohazards, raw materials, mineral resources, marine environment, water resources and soil. The

EGDI is based on the data and expertise from collaborating geological survey organisations in Europe and will be managed under their responsibility.

The EGDI will support the continuous maintenance and development of relevant INSPIRE-specifications and extensions, portals, high-level software tools, web services and connections to external data infrastructures.

Main objective is to enable communities – policy departments, companies, researchers and geological surveys - to provide geological services at cross-border, European and international levels. These services involve areas such as policy support, hazard management, industrial development and environmental monitoring.

EGDI must consider various types of actors:
- Data and services providers (mainly the Geological Surveys)
- Users of EGDI resources (professional users able to process the geological information)
- Users of geological products (final users, decision makers)
- Technical engineers able to manage the system

The design of the EGDI architecture must be compliant with INSPIRE requirements as the goal is to deliver information related to several INSPIRE data themes as geology, hydrogeology, mineral resources, natural risk zones, energy resources. The INSPIRE requirements cover the architecture (distributed system based on services), the metadata, the data specifications (data models, code-lists), the web services (at a minimum discovery, view and download services, but also spatial data services), the data sharing.

The main pattern of the EGDI architecture is a distributed information system (as the data and services providers are the European Geological Surveys) and services oriented (often named SOA - Service Oriented Architecture).

INSPIRE and GEOSS are examples of such architectures, but also OneGeology-Europe.

The OneGeology (global) Spatial Data Infrastructure formally provides a 'core global dataset' to GEOSS.

As a distributed system, EGDI must request the implementation of data interoperability arrangements (as defined in INSPIRE) among all partners. They are the rules for collecting, storing, processing and disseminating shared data, metadata and products.

Being THE European counter for serving geological data, EGDI must be connected:
- to International Initiatives such as GEOSS / OneGeology,
- to Research Infrastructures (such as EPOS) to provide geological information to researchers and vice versa,
- to supported projects that runs under EGDI umbrella.

Figure below presents a possible architecture.
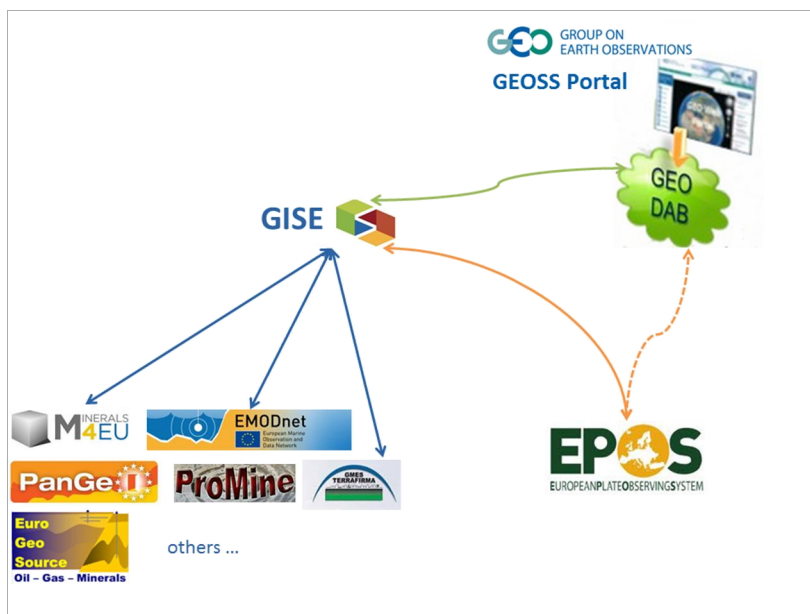
Figure 2: Possible interaction between EGDI-EPOS-GEOSS

Interdisciplinary interactions should be set up with various domains (ex: Surface water, Marine, Risks, …). In order to deliver better information to those domains it is strongly advised to apply INSPIRE requirements and continue applying those when extending data models, Code-Lists and making evolution to software tools.

EGDI should also provide THE contact point to those domains maintenance secretariat (ex: WISE Technical Group for Water) in order to ensure that their needs are also taken into account and both semantic and technical interoperability are working on cross-domain either.

Data sharing within EGDI will be fostered when getting as close as possible to open and documented products. Datasets should be (when feasible) available free of charge or at minimum costs.

Eventually, in order to properly support NGSO communities in their IT deployments, EGDI will provide a structure to exchange and test the latest technologies available. This, to assess their real added value to the community

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:
Requirements #1, 2, 6

4. Information Viewpoint

This viewpoint should focus on the semantics of the information and information processing performed, by describing the structure and content types of supporting data.

4.1. Metadata

The metadata is the structure describing EGDI resources. They are used by the providers to register their resources in the EGDI catalogue and by the users to discover relevant resources (datasets, services, documents, best practices …).

The recommendation is to use INSPIRE metadata elements (based on ISO 19115/19119 standards).

As EGDI products or datasets will be an aggregation of national data, a metadata record must be available for each European product or dataset, and the provenance must be specified (possibly with the lineage element of the metadata standard which addresses the sources and the production processes used to create the resource). A link to the metadata related to each dataset used for the European product must be available.

For each layer the metadata must be available through the EGDI portals (click a "I" or "?" button). All metadata are stored in a catalogue and made available, and searchable, by a catalogue service (a main component of EGDI architecture, see Service viewpoint section)

The metadata will give access to the services URL (to be displayed in the portals, as a property of a layer)

Some identified keywords should be used to make the discovery easier (as geological map of Europe, geology Europe, name or acronym of the European project that generated a given dataset …)

Multilingual issues will be taken into account. For example using multilingual thesauri

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 5, 16, 17, 18, 19, 20, 21, 25, 39, 42, 63

4.2. Datasets priority

WP3 has defined the priority for datasets that must be taken into account:
- Geology
- Mineral Resources
- Water Resources
- Geohazards: flooding, earthquakes, landslides, subsidence
- Soil

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 43, 44, 45, 46, 47, 48, 49, 50

4.3. Data models

WP4 – D 4.2 advised the following rationale:
- Use INSPIRE Data Specifications, when exist, as a basis, maybe use extensions if needed (based on international standards when they exist),
- Use also the results of data model implementation in projects, such as the Minerals4EU project for implementing Mineral Resources,

It also identified the following list as a starting point for EGDI (see D 4.2 for more details):
- Inspire
  - Generic Conceptual Model,
  - Geology,
  - Mineral Resources,
  - Environmental Monitoring Facilities,
  - Area management/restriction/regulation zones and reporting units,
  - Hydrology,
  - Soil,
  - Natural Risk Zones,
  - Energy Resources.
- CGI
  - GeoSciML
  - GeoSciML-Portrayal
  - EarthResourceML
- OGC :
  - WaterML2.0, GroundWaterML2.0

As presented in D 3.3, methodologies will be defined to generate derived datasets (EGDI products). In some cases there might be need to clearly differentiate models for "raw" data from product specifications.

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 43-50, 51-53, 55, 57, 59, 61, 62, 68

Data models for exportation to GIS tools

Often GIS tools cannot manage complex data models, so there is a need to define a simple feature data model to deliver data to these tools. It is the recommended to define such simple feature data models according to possible standards.

The CGI has specified such data models (in the standards GeoSciML-Portrayal and EarthResourceML), named Portrayal Classes, which group several properties of several classes into one class linked to one geometry (point, line, or polygon).

Links to user needs and functional requirements (WP2 - D2.4)

The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 28

Data model related to 3D Geological models

Up to now, the data of 3D models are stored according to the structure of each modeling tool (GOCAD, GSI3D, GeoModeller, …)

4.4. Maintenance of data models

EGDI must define governance rules to address the maintenance of the data structure and the versioning management.

In order to achieve maximum interoperability, this governance should also take into account close interaction with the structure(s) responsible of the maintenance of the identified relevant data models (ex : Inspire, CGI, OGC, …).

Links to user needs and functional requirements (WP2 - D2.4)

The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

4.5. Code-lists/vocabularies

Code-lists/vocabularies should be defined, maintained and made available according to INSPIRE rules. For example, clear rules are defined for INSPIRE code-lists that are extensible.

INSPIRE rules for encoding (SKOS/RDF) and for identifier (URI) should also be applied.

Shared code-lists/vocabularies might also be translated in a coordinated way in order to help solve multilingual issues.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 54, 56

4.6. Maintenance of code-lists/vocabularies

EGDI must define governance rules to address the maintenance of the content (concepts, terms), the multilingualism, and the versioning management.
Some tools are necessary to manage the collaboration (providing comments, new concepts and terms) among experts; and web services to deliver, to access and to use vocabularies.
For the same reason as for data models, this governance should also take into account close interaction with the structure(s) responsible of the maintenance of the identified relevant data models vocabularies (ex : Inspire, CGI, OGC, …).

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

4.7. Integration of heterogeneous data

EGDI must also be able to integrated heterogeneous data coming from multiple sources.
While data models and shared code-lists/vocabularies are the core objective to ensure proper semantic interoperability, there need to be a possibility to also take into account data that are not "following the rule". This could happen for various reasons (ex: provider outside EGDI umbrella, data from a former project with no expertise available anymore, …). To solve those issues, work done on semantics, ontologies must be taken into account.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

4.8. European products

EGDI has to consider many (national) datasets or maps as **one** European dataset or **one** European map of "something":
- To describe by one metadata record (including the link to the metadata of each national dataset or map processed to create the European product)
- To use the same legend, and then the same portrayals rules

- To describe the methodology used to create the national dataset or map

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

4.9. EGDI Registry

This registry and its content is a **key component** for EGDI: it should contain (publically available or not) common elements to be shared among geological surveys from scientific concepts, best practices, vocabularies, … to very technical elements for implementation (standards to use, software components, …)

As a first proposal it could contains list, information, links, … about:
- Components: List of available components (portals, catalogues, software, …) from projects
- Standards: List of standards recommended by EGDI for a project development
- Best practices: Best practices to help implementations
- Vocabularies: Existing vocabularies (code-lists) and related web services
- Objects dictionary: Definition of objects managed in projects (= Feature catalogue)
- Portrayal rules: Common rules to portray features, and Style files
- …

This list should be flexible and updated based on feedback received from projects managed under the EGDI umbrella.

Detailed description of registry items:

Components:
List of available components (portals, catalogues, software, …) from projects. This list should provide:
- the web address of all portals available (who knows that to get information about mineral resources one has to use ProMine or EuroGeoSource portals?)
- the list of the applications developed
- the catalogues available to be connected to new "geological" portals if necessary or to searched by other portals,
- software to create a portal, to setup web services, to process data, …

Standards:
List of standards recommended by EGDI for a project development about:
- Metadata (with the link to INSPIRE requirements)
- Web services: discovery, view, access, process, …
- Data models: GeoSciML, EarthResourceML, GroundwaterML, QuakeML, …+ INSPIRE data specifications

- Data quality recommendations

- *…*

Best practices:
Best practices related to methodologies ("how to make a radon map") or technical implementations ("how to setup a WMS for Mineral Resources data with MapServer")
This section could be organised according to thematic domains.

Vocabularies:
Existing vocabularies (code-lists) and related web services: it is known that it is very time consuming (and expensive) to define common vocabularies. Also they should be well managed at European level, so that they can be visible, re-used and improved. Good examples are:

- the GEMET thesaurus about environment (http://www.eionet.europa.eu/gemet/)
- the INSPIRE registry (http://inspire.ec.europa.eu/codelist/)

Objects dictionary:
Definition of objects managed in projects (= Feature catalogue). As for the vocabularies, the definition of geological objects should be visible, re-used and improved at European level.
As an example the INSPIRE Feature Catalogue: http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY/items/26

Portrayal rules:
Common rules to portray features, and Style files. To move to more interoperability a first step should be to use the same portrayal rules. As to agree on them is time consuming and also to implement them (in a WMS for example) EGDI should offer to share them and provide the style files.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

4.10. Collaboration with international organisations

To maintain the EGDI data models and code-lists it is recommended to establish relationships with international organisations as EC-JRC (about INSPIRE), IUGS/CGI, OGC, and others.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

4.11. Data policy, licenses

User registration could be setup in EGDI to control the access to the data but also to record information regarding the use of the data.
WP5 has detailed those aspects. Some decision might have a direct impact on EGDI architecture.
INSPIRE maintenance actually considers work done in other European projects on Authentication, Authorisation and Accounting.
This could provide important guidance both on the organisational and technical aspects

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

5. Service Viewpoint

This service-oriented viewpoint enables distribution through functional decomposition of the system into objects that interact at interfaces. It is the core viewpoint for the identification of interfaces and services.

EGDI architecture will be a Service Oriented Architecture (SOA), a distribution through functional decomposition of the system into components that interact at interfaces over network communication services.
At a minimum, the functions will be those requested by INSPIRE for national infrastructures where spatial datasets, described by metadata, are served by services for discovery, view, download, and invoke (processing services).
As much as possible, all EGDI services must be compliant with INSPIRE specifications, and OGC Specifications when exist.
This section describes a list of services to be implemented by EGDI. The next section (the Engineering viewpoint) identifies the components with the standards to use, when exist.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 26, 36

5.1. Discovery services

The EGDI catalogue is a "classical" catalogue already implemented in several projects, registering metadata of available datasets and services.
It implements standard specifications so that it is also accessed by other portals (as GE-OSS portal). This catalogue is connected to a search engine available in the portal.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 2, 4, 8, 18, 19, 20, 22, 23, 24, 58

5.2. Map services

The map service is for accessing and visualising data, and must:
- give access to:
  - o the legend related to the map displayed;
  - o the information related to the object (point, line, polygon) selected by the user (get feature info operation of the map service)
  - o several projections
- support a zoom dependant of the scale level (to display objects only at a certain level of zoom)
- be able to display gridded data (connection with a Web Coverage Service)

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 4, 5, 8, 12, 13, 14, 15, 25, 26, 27, 59, 65, 69

5.3. Data access and sensor web services

Data access
The data access function is a service to provide some data to an application that is able to process these data to deliver a result. It is not seen as a download service described in the next section, even if the software tools could be the same.
In the interoperable world implemented according to OGC standards and INSPIRE rules the data are delivered in a GML application schema (XML stream), which is processed by an application. When available, the GML schema must be compliant to the INSPIRE Data Specifications.

<u>Sensor web services</u>
They host services to describe and access sensors and sensors networks, and to access sensor data. Other services related to sensor planning are probably out of scope of EGDI. The sensor data are described in a standard way using the observations and measurements specifications and, most of the time, encoding.

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 25, 26, 27

5.4. Download services

The download function is not seen as the access data function described in the previous section, even if the software tools could be the same. For the download, the user wants to get a file on his computer to process the data with a dedicated tool.
If the data format for the access data services is a GML application, compliant to the IN-SPIRE Data Specifications (as requested by INSPIRE), the download service must offer various other possible formats (ArcGIS, MapInfo, spread sheet, NetCDF, …)

Files delivered by a download service must be a package containing:
- The data in the requested format (shapefile, spreadsheeets, GML, gridded data – Ascii grids, NetCDF…)
- The related metadata, containing the licence (copyright, conditions of use)
- The legend (so that the data can be portrayed easily without recreating the legend)

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 25, 26, 27, 28, 29, 30, 31, 32, 36

5.5. Processing services

Processing services is a generic term to name a group of services offering many functions to be specified. Examples of such services could be:
- A thematic analysis service (developed for the OneGeology-Europe project) able to display a geological map with only geological units of a certain lithology
- A statistical service able to compute the number of geologic units of each type of lithology within a specified bounding-box (it could be also the area of each unit).
- A graph service delivering dynamically a graph of the ground motion time series related to a PSI measuring point (geohazards requirement – from PanGeo)

- A report service able to deliver a report containing geohazard descriptions of the individual PanGeo towns

They host services to process data accessed from a remote service (WFS, WCS, …). They can be generic (compute statistics for example) or dedicated to a thematic domain.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 8, 25, 26, 33, 34, 66, 67

5.6. Gazetteer service

This service is for searching a name of a location, and having the map automatically zooming to the desired location. This service is used by the portals.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 63

5.7. Services related to 3D models

This service will allow sharing content related to 3D models (Virtual borehole, Sections,…). Of course 3D viewer will also be needed.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

5.8. Model Web services

These services allow to access and visualize results from simulation and modeling tools.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

5.9. Brokers

Brokers (also termed mediators) are software components to discover and access data and services as probably all these resources will not implement the same standard.

To re-use data from previous projects or to connect other infrastructures which do not use the same standards, it might be necessary to develop some software components ("brokers") on top of existing resources.
These brokers could be for:
- Accessing metadata not implemented with the ISO standard,
- Accessing data delivered according to various formats,
- Harmonising coordinates of various sources of data,
- Providing terms (from vocabularies) in one language,
- Delivering data into formats requested by thematic tools (GIS, 3D Modeling, …)

Brokers already proved their value added by lowering the entry barriers to both Data Producers' and Users in international initiatives such as GEOSS. In GEOSS, the Discovery and Access Broker deployed is the cornerstone of the system.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

5.10. Workflow management

This component is an engine able to manage workflows to automate tasks and to combine processing services.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

5.11. Data validation - Conformity testing

The European datasets or maps are the result of the aggregation of national datasets or national products. The "quality controller" defined in the EGDI governance structure should have some tools to check the conformity of the data to EGDI Specifications.

The web services deployed according to EGDI specifications should be tested against those specifications. This will support the deployment of national datasets and, in turn, facilitates the generation of European datasets and maps.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39, 40, 41

5.12. Access control, security, user management

The operational group responsible for the technical maintenance of EGDI should be able to have the necessary credentials to do so.

For some European datasets and maps, access control levels could also be taken into account.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

5.13. Quality of Service (QoS)

In order to properly achieve its role of central repository for Geological data in Europe, EGDI portal and associated services should reach a certain Quality of Service
Being compliant with INSPIRE requirements, EGDI infrastructure should at least respect Quality of Service requirements defined in the COMMISSION REGULATION (EU) No 1088/2010 of 23 November 2010 amending Regulation (EC) No 976/2009 as regards download services and transformation services (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0001:0010:EN:PDF). More information including associated Guidance Documents are available on the INSPIRE website (http://inspire.ec.europa.eu/index.cfm/pageid/5)

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 10, 11, 39

5.14. Monitoring of infrastructure services

Once deployed, the services infrastructure should be monitored in order to ensure the allocation of hardware resources corresponds to the use of the services.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 11, 39

5.15. Ontologies services

In order to be able to integrate data from multiple sources, ontologies should be considered by the EGDI not only for vocabularies but also for data models and their corresponding instances.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

5.16. Maintenance of service specifications

EGDI must define governance rules to address the maintenance of service specifications. It is recommended to establish relationships with OGC to collaborate to the service specifications process.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

6. Engineering Viewpoint

This viewpoint Identifies types of component to support distributed interaction between the components

6.1. Overview of the EGDI Architecture

The objective is to identify a generic framework in which existing and future projects could be described, and which guarantees that EGDI will be able to communicate with or to have some EGDI components as a component of other infrastructures (as EPOS for example).

The general motivations are:
For the providers:

- To manage how and when data are delivered to the users,
- To use existing interoperable web services from other thematic domains necessary to display or process its data, without having to download and manage these "external" data. For example topographic maps or hydrographical maps can be displayed directly using Web Map Services with no need to download data.
- To have its data available not only for some projects well identified but also for other usages not yet known

For the users:

- To access via portals and applications various resources in an interoperable way, with no need to develop specific interfaces for each provider
- To understand data provided by many providers as they share a common data model and dictionary

The EGDI architecture is based on a "3 layers" architecture:

- An **access layer**: it contains the data services produced by the geological surveys at the national or regional level. Most have to be INSPIRE compliant and are implemented through WMS, WFS or WCS web services (other download services could be added later on).
- A **mediation layer**: it contains the common components that are required to register, view, access and process data. An example of such component is the catalogue that identifies all the elementary data services from the access layer. Some "standard" processes can also be identified in this layer, such as interpolation services (to interpolate a map from point data) or statistical tools (to produce pie charts from selected data sets).
- A **client layer**: it is the "visible" component of the "portal", and contains the EGDI portal, thematic portals, or smartphone apps for instance. It uses services delivered by the mediation layer or by the access layer.
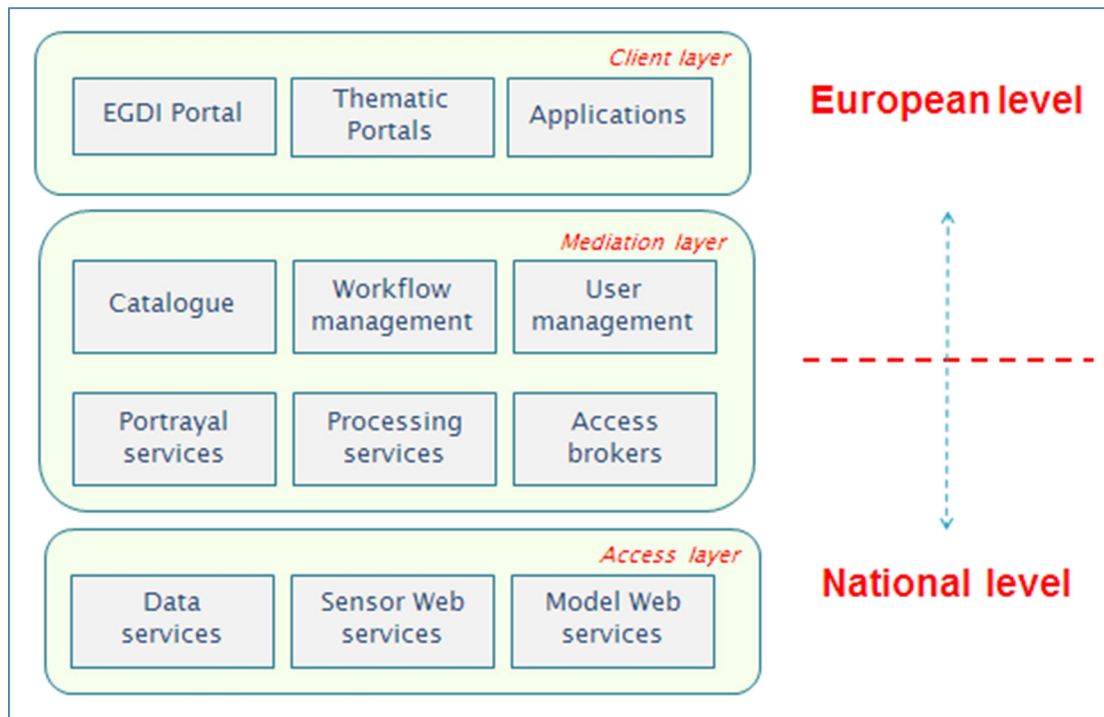
Figure 3: Overview of the framework proposal for the EGDI architecture

The architecture must also manage **common elements** shared by many projects and applications:

- A **catalogue** of metadata about datasets and services available
- And a **registry** of:
  - Components, portals,
  - Standards
  - Best practices
  - Vocabularies
  - Objects dictionary
  - Portrayal rules
  - …
- And a set of software components (**brokers**) to discover and access various data and services as probably all these resources will not implement the same standard.

All these common elements must be visible and accessible from a specific portal, the **EGDI portal.** This resources portal is not especially for "final" users but for Geological Surveys or experts interested by using existing resources. This portal is the presentation of the catalogue of all available resources (in a broader sense), associated to a search engine and a viewer.
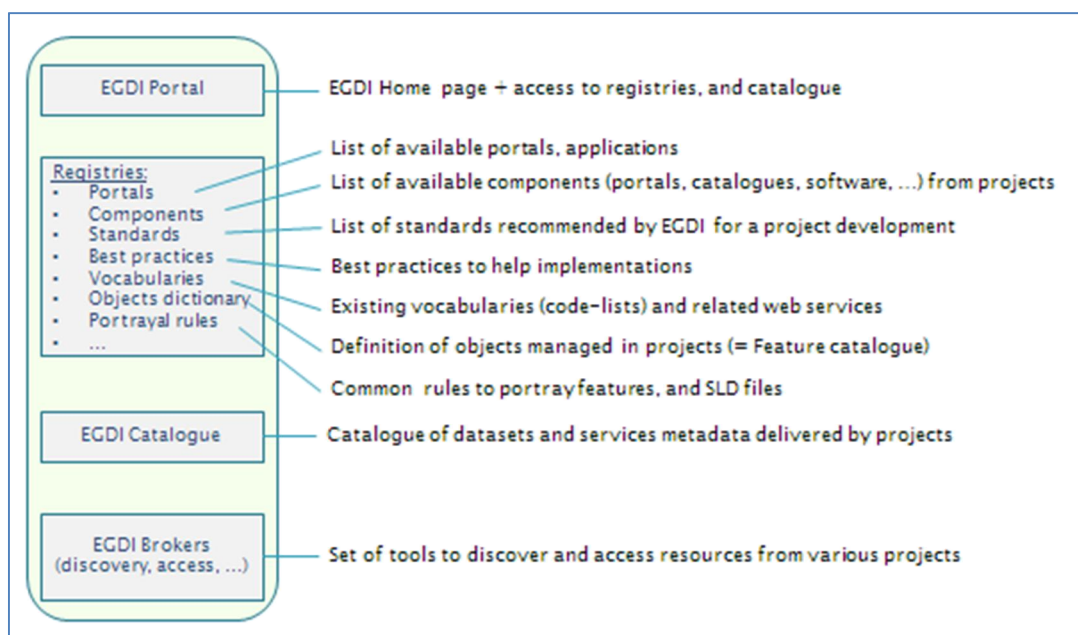
Figure 4: Content of the EGDI Portal

6.2. Architecture: EGDI Portal

The use case for this portal is to give access to all common resources. It is not a portal for "final" users but for geoscience experts able to use and combine resources to create a European product.

Entry point of the Portal, the **EGDI Viewer** will enable user to interact with dataset/products aggregated at the portal level.

It will also provide access to:
- The **EGDI Catalogue** with a search tool that will display:
  - The metadata of datasets and services (mainly the view and download/access services)
  - The metadata of processing services (all services able to process data)

From the datasets metadata, it should be possible to display data (in a very simple way), and to get the data description (more than the metadata which describes the whole dataset; it is here to describe the objects and their properties like a feature catalogue), and possibly to download data.
- The **EGDI Registry** managing a list of "components": projects and their portal, dedicated to a thematic domain, standards used, best practices, vocabularies, …(see below the description of the EGDI registry content)

The OneGeology-Europe portal could be a first example: it gives access to map layers developed by various projects, but it is not dedicated to specific users (final users).

Usually portals developed for several projects offer the following functions:

- To discover datasets described by metadata registered in a specific catalogue (more or less standard)
- To view the data, often thanks to a WMS
- To download (thanks to a "ftp like" loader) or access data thanks to a WFS
- To process data (statistics, charts, graphs, …)

The design of those portals has to implement a specific branding (graphic charter) defined by the project.

It seems difficult to include all these functions into **one** EGDI portal:

- The metadata, the catalogue and the data viewer could be at the European level into EGDI
- But for the specific functions, based on specific data models, it could be difficult to find a user friendly way to present them to a final generic user. This is the role of dedicated portals that must be designed targeting users identified for a given project.
- The specific branding (graphic charter), that could be requested for a category of users or mandatory for some projects, cannot be mixed with others

Then it is not possible to design one EGDI portal for all kind of users. But there is a strong need to have a portal providing access to all common resources to be shared for running and future projects. This portal is for Geological Surveys experts involved in various projects.

The same discussion is also valid for applications and smartphones applications: probably there is a need to agree on some common components but the mobile apps are very specific to a group of users.

An adapted OneGeology-Europe portal (graphic charter, updated / new functionalities) could be considered a first version of an EGDI portal. It provides access to datasets through services described by their metadata registered in a catalogue. For several projects (OneGeology, Pangeo, ProMine, EMODnet, …) the main services initially offered are the discovery and the view services. OneGeology-Europe also provides some specific services (to make a map of selected geologic units according to lithology and/or age, and statistics on an area).

Links to user needs and functional requirements (WP2 - D2.4)

The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 2, 3, 4, 6, 8, 9

6.3. Architecture: EGDI Catalogue

This EGDI catalogue is a "classic" OGC/ISO catalogue already implemented in several projects, registering metadata of available datasets and services. The registered metadata are compliant with the INSPIRE metadata regulation (based on ISO 19115/19119 standards).
This catalogue is connected to a search engine available in the portal. It implements the standard specification CSW from OGC. The catalogue could be accessed by other portals (as with the GEOSS portal).

The catalogue being the cornerstone of the system and given EGDI objectives, highest interoperability should be achieved with other initiatives. Thus EGDI catalogue should also provide OpenSearch endpoints and bridges with both CERIF and DCAT. The latter two will be made possible as mappings from INSPIRE metadata to both CERIF and DCAT-AP are ongoing.

Search engine to be implemented (ex :Solr, elastic search, …) will also be crucial as it does the interaction between the user and the EGDI catalogue.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 5

6.4. Architecture: EGDI Registry

EGDI registry will be the place we all shared vocabularies, reference datasets, feature catalogues elements will be accessible.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

6.5. Architecture: EGDI Thematic Portals

To address thematic user requirements it is necessary to develop dedicated portals that could be stand-alone portals using EGDI services, or various tabs in the EGDI portal.
Apart from the graphical interfaces aspects mentioned above, thematic portals will be exchanging data with the EGDI Portal.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 3, 6, 7, 8, 9

6.6. Architecture: Data collection

Applying INSPIRE principles, data will be collected from National levels using the Service Oriented Architecture deployed.

Figure 5: INSPIRE architecture principles

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 36

6.7. Architecture: European database vs National databases

Two main aspects can lead to the set up of an EU level database:
- Without such a database, data disseminated by the EU portal/web services must be validated on the fly which can be a risky exercise. Issues detected requiring sometimes exchanges with the national data provider. This could lead to establish a database at the EU level; thus at the EGDI registry level,
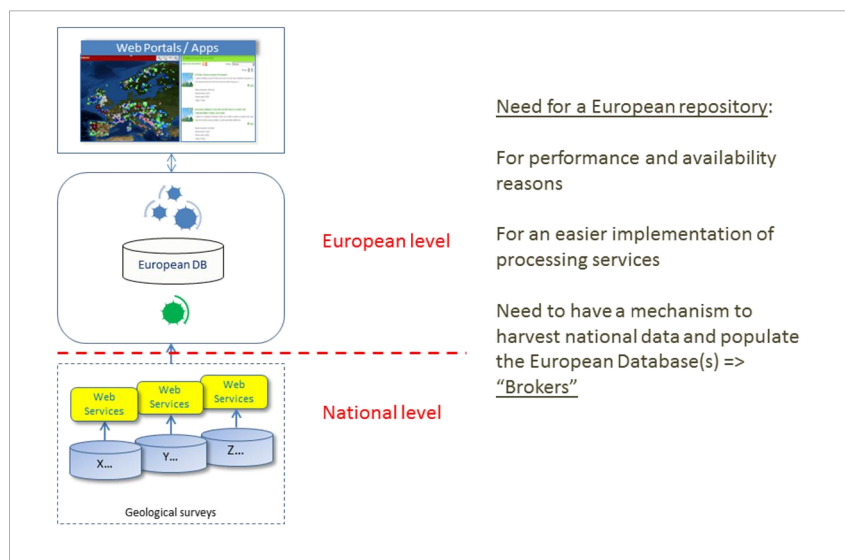
Figure 6: European database needs

- To enable EU products generation process as described in the figure below
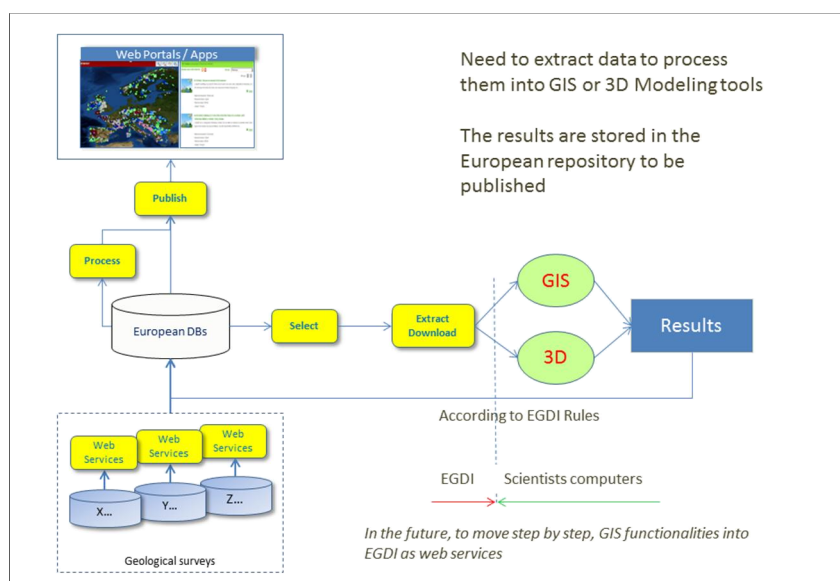
Figure 7: European database and EU products generation

When such a European database is setup it must be managed and maintained. The various tasks should be taken into account:
- Setup the central database(s)
- Create the connections with the national data providers.
- Manage the mechanism harvesting the national data to fill the European database(s). Various techniques can be deployed: ETL tools, webservice transfer, …
- Manage the changes in the data model when necessary (incremental change, full refresh, …)

The EGDI governance must decide how the tasks described above are taken into account.
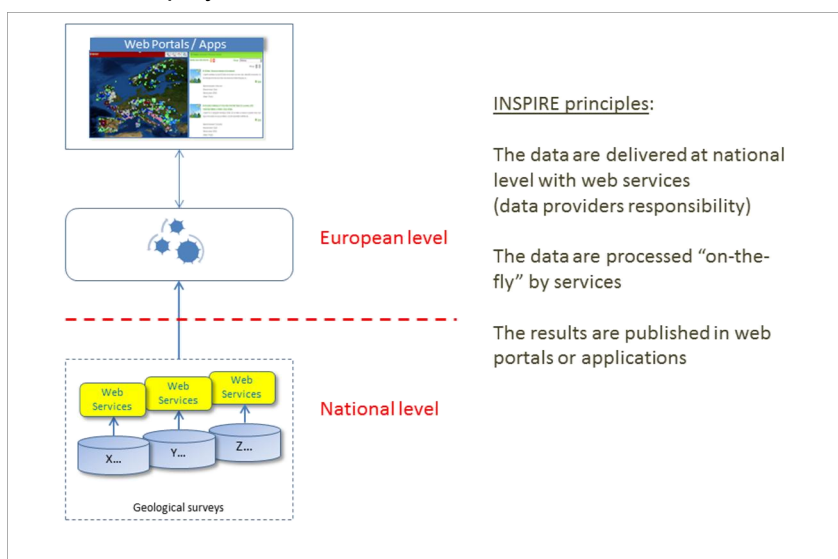
<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 38, 39

6.8. Architecture: Harvesting database vs Dissemination database

In case a European database is to be set up another aspect to be determined is the need to have at the EU level:
- A Harvesting database fed by national databases (see former chapter),
- and a Dissemination database synchronised with the harvesting one and only for data dissemination purposes. This database being the information source for the web portals and data dissemination webservices.

The interest of such an architecture are multiple:
- The harvesting database can be updated without stopping the data dissemination flow as sometime harvesting national information can be time consuming,
- The harvesting database can ensure data quality/conformity before data dissemination,
- The dissemination database can be fine tuned so that the Quality of Service is ensured.
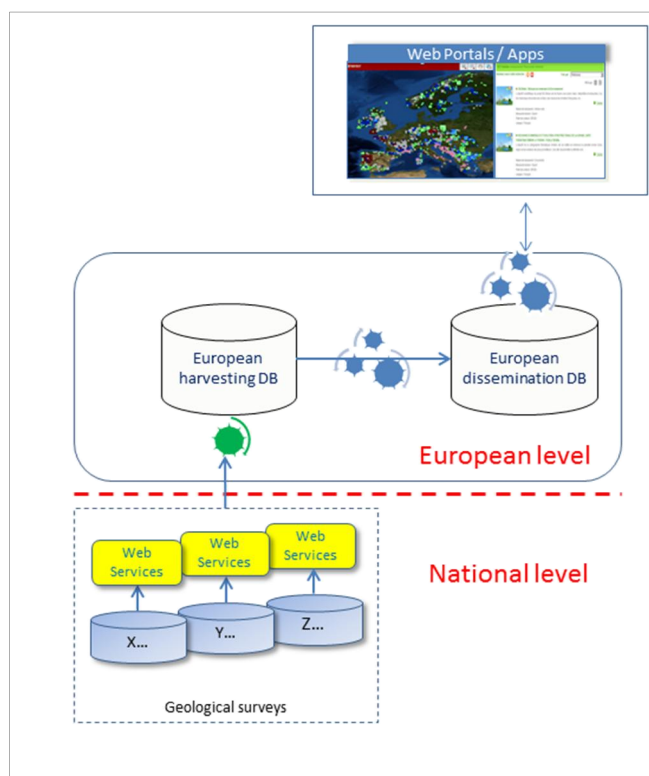

Figure 8: EU Harvesting/dissemination databases

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39, 40, 41

6.9. Service architecture standards

To ensure services interoperability, it is recommended to follow the OGC Framework.
Standards specifying the services described in the Service Viewpoint exist. These are indeed most of the time already available at OGC level.
If EGDI wants to share these services in the INSPIRE community, it is recommended to follow the Spatial Data Services regulation (based on the WPS for the best defined interface of the possible INSPIRE Spatial Data Services types).
It is recommended to follow specifications mentioned below.

- For discovery services, OGC CSW (Catalog Service for the Web) standard is recommended.
- Map services deployment concerns various OGC standards.
    o WMS (Web Map Service) and WMS-T (WMS + Time)
    o SLD (Styled Layer Descriptor) and SE (Symbology Encoding)
- Data access services emcompass
    o WFS (Web Feature Service)
    o WCS (Web Coverage Service)
    o SOS (Sensor Observation Service)
    o Other standards can be added to that list (ex : TJS – Table Joining Service) as various solution coexist depending on the use cases)
- Download services often concern the same services as data access services. FTP solutions can also be taken into account once volumes to be exchanged grow bigger,
- Processing services standards are
    o WPS standard (Web Processing Service)
    o and WCPS (Web Coverage Processing Service) to process gridded data.
- Gazetteer service functions can point to two different standards within OGC
    o WFS-G : a gazetteer profile of WFS
    o Or OpenLS (Location Service)

So far, there are no OGC endorsed standards enabling 3D model output sharing. Various initiatives already exist. Some could be followed and tested by EGDI.

On the same line, there are no standard Model Web Services specifications. Joint initiatives between EGDI and the OpenMI initiative could be way forward to solve that issue.

Eventually, RDF endpoints should be considered as ontologies services. The on-going discussions under INSPIRE maintenance process should be followed.

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 2, 4, 5, 8, 12, 13, 14, 15, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 58, 59, 63, 65, 66, 67, 69

6.10. Brokers

There are no existing standards on Brokers (mediators). But already proven solutions do exist (ex : GEOSS Discovery and Access Broker).
When the need to implement Brokers at EGDI arises, it will important to decide between going for an ad-hoc solution or try an investigate Brokers standardisation.

<u>Links to user needs and functional requirements (WP2 - D2.4)</u>
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

6.11. Data validation – Conformity testing

Being compliant to INSPIRE recommendations and based on INSPIRE practices, main data exchange will be based on XML encoded data flow.
Basic conformity testing could then naturally be carried out checking the XML content validity against the XSD file generated from the UML data model.
More in-depth checking could involve Schematron rules (comparing various attributes values, testing threshold).
When testing geographical aspects will be required (ex : distance between objects, object shape conformity, …) other technologies will then be required (ex: loading datasets into a spatially enabled database).

Requirements # 39, 40, 41

6.12. Access control, security, user management

Important European projects involving national security agencies already work on providing guidance on Authentication, Authorisation and Accounting.

User registration and SSO (Single sign-on) by using an ID Service: use of Open ID (http://openid.net/) or other specifications and tools (such as Shibboleth - http://shibboleth.net/ ; SAML2 - http://en.wikipedia.org/wiki/SAML_2.0 and GeoXACML-http://en.wikipedia.org/wiki/GeoXACML) are being discussed and benchmarked.

As this is done by the INSPIRE maintenance process, considering the work done by those projects will provide important technical guidance.

6.13. Maintenance of components

Each of the components deployed should also me maintained both from a thematic and technical point of view.
This concerns:
- Maintenance of the shared data models and codelists (vocabularies),
- Maintenance of discovery portal,
- Maintenance of thematic portals,
- Maintenance of database.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 39

7. Technology Viewpoint

This viewpoint identifies component instances as physical deployed technology solutions, including network descriptions.

7.1. Supported platforms

EGDI Resources portal and related services must be available on the web and on tablet computers.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements # 9

7.2. Service architecture deployment

There are various solutions compliant to the different OGC standards identified. Both open-source and vendor solutions compliant to GOC standards are listed on the following link: http://www.opengeospatial.org/resource/products/compliant

7.3. Tools to help data publication

EGDI, through various projects running under its umbrella should share tools developed by EGDI to help data providers to publish their data accordingly to EGDI rules.

In order to ease sharing of such tools, it will important to take a close look at tools licenses. Open Source solutions (depending also on the open source license) might help in that respect.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

7.4. Cloud technology

The previous chapter advises tools sharing under EGDI.
But not all NGSO have always the required IT knowledge available at hand. Moreover depending on the tools shared, the technology required might not be the same.
Another solution could be to set up the necessary tools in a Cloud based solution which will remove tool installation issues in all NGSO.
Such solution has been studied in the InGeoCloudS CIP project.
This could lead to an evolution in the architecture with NGSO willing to deploy shared tools in their own infrastructure and others using the same tools but already deployed in a Cloud based one.
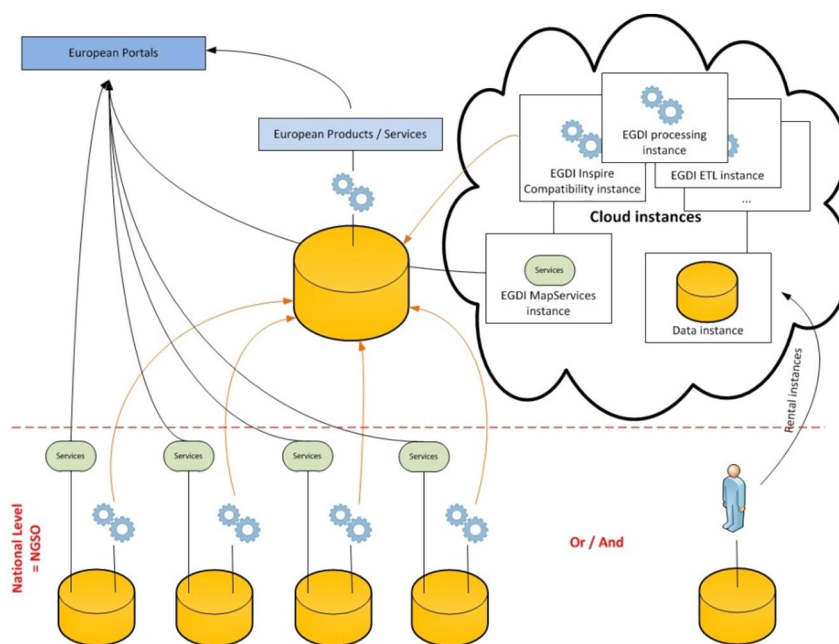
Figure 9: Mixed infrastructure including Cloud

Links to user needs and functional requirements (WP2 - D2.4)

The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

7.5. Hardware infrastructure requirements

The concrete deployment of the architecture, standards and tools solutions described above will have to be decided when an implementation is possible.

The given context of that implementation will lead to delineate more precisely the architecture, standards and tools needed.

Depending on the context and choices made, hardware requirements to fulfill the Quality of Service identified above could change drastically.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

7.6. Maintenance of software/hardware

Soft/hardware deployed for the EGDI Portal and also EGDI Thematic Portals will also have to be maintained. Sharing common tools/practices will reduce those constraints.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

7.7. New coming technologies

EGDI must ensure permanent technology watch to ensure that new coming technologies added value is properly evaluated under EGDI. This will, in turn, provide NGSO and European projects with proper IT feedback on technologies such as heterogeneous data integration, ontologies, data mining to name a few.

Links to user needs and functional requirements (WP2 - D2.4)
The architecture needs described in this section come from the following user needs and functional requirements identified in the WP2 - D2.4 document:

Requirements #

8. Skills and training needs for individual NGSO

Each individual NGSO might implement and/or contribute to the proposed EGDI system described above. Depending on each survey capacity, training might then be needed.

Improving the individual capacities of the NGSO members as well as their contribution to such a distributed SDI within a continent-wide network and collaborative initiatives should be associated with a dedicated training program built on the personnel and skills already in operation.

Identification of equivalent training needs for individual NGSO was carried during the FP7 AEGOS project[1]. The rationale is here adapted to the EGDI context.

The detailed assessment carried out as part of the AEGOS design phase showed that not only a solid knowledge and practice basis in producing digital information but also advanced skills in data security, data modelling, quality assurance, complex data analysis, data storage and controlled publication were needed. Complementing this analysis and based on the output of the current EGDI-Scope exercise, the following roles can be added: Data provider, Publisher, Discoverer, Service provider, Service orchestrator.

Three target groups were identified including IT personnel (infrastructure, administration and interoperability), geoscientists (information management and processing) and GIS professionals (user-oriented products and services).

The knowledge base existing in the surveys should be reused as much as possible and the training scheme should be carried out through three levels i.e. basic, foundation and expert. (see Figure 10 below)
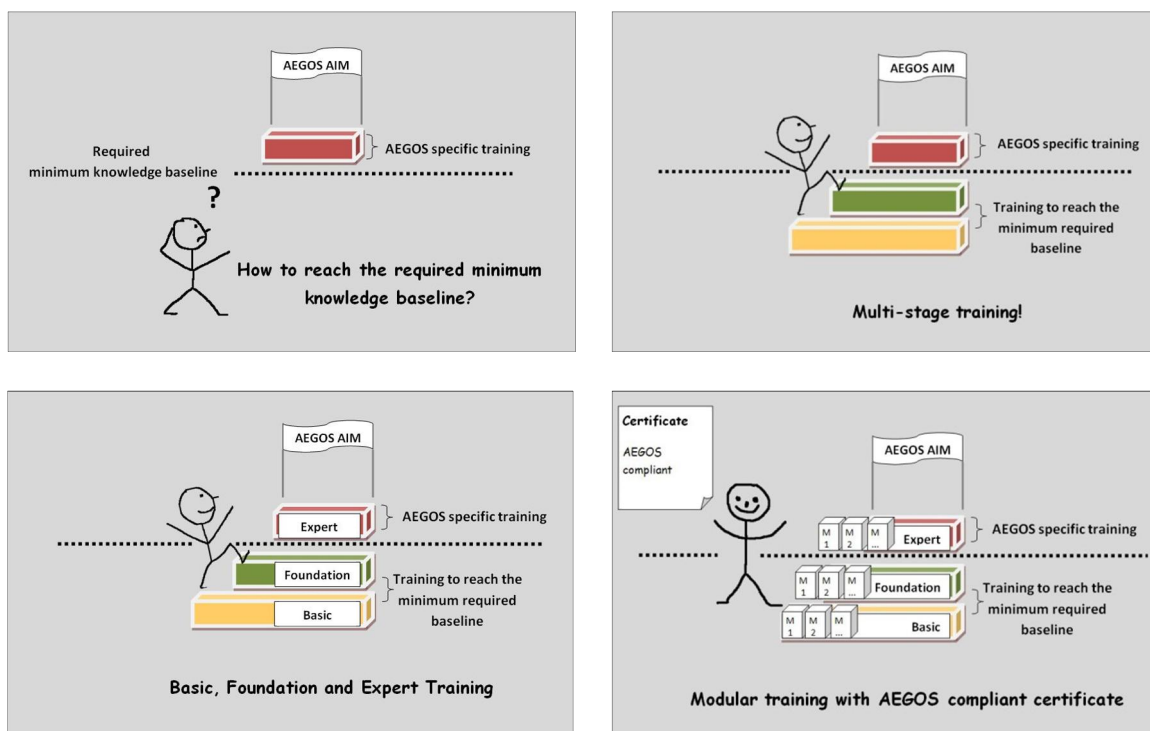

Figure 10: Multi-stage training defined in AEGOS

The AEGOS capacity building program proposed a set of 28 modules to be implemented through regional and/or local sessions (an EGDI adapted example is proposed at the end of this chapter). The training methodology included face-to-face, on-the-job and open distance electronic learning training sessions, info days and workshops. This scheme included

[1] European Commission FP7 – Project no. 212545 – African-European Georesources Observation System

training for IT and geodata professionals as well as training for trainers in order to antici-pate the staff turnover.

European NGSO already have a long lasting cooperation experience in shared projects (ex : EU FPx, CIP, eContentPlus …). This cooperation along with the activity generated by INSPIRE directive is a unique context for knowledge transfer.
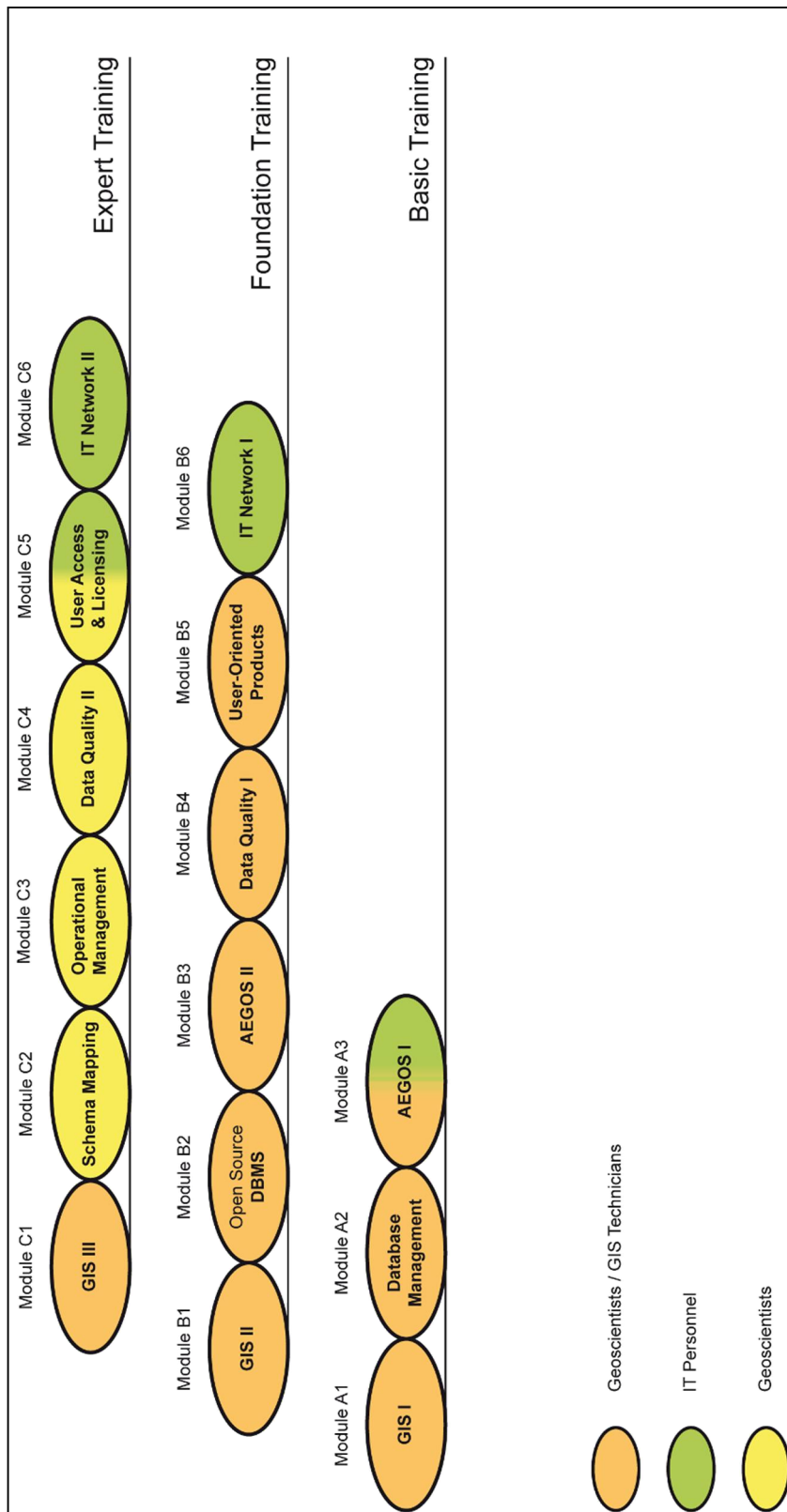
Figure 11: Modules of the multi-stage training in AEGOS

Module	Topic	Methodology
BASIC Training		
GIS I	From analogue to digital maps	OJT
	Introduction to Desktop GIS	FTF & OJT
Database Management	Introduction to Database Management	FTF
EGDI I	Introduction to EGDI	ID
FOUNDATION Training		
GIS II	Advanced methodologies for data analysis and data processing	OJT
Open Source DBMS	PostgreSQL and PostGIS	FTF & OJT
EGDI II	General introduction to the principles of SDI	S
	General information about INSPIRE, EGDI, GEOSS, OneGeology	ID
	Practical Exercises in using external WMS and WFS	S
Data Quality I	Data Quality Assurance for Technicians	FTF & OJT
	Data Quality Assurance for Project Manager	FTF & OJT
User-oriented Products	How to create value-added products	FTF
IT-Network I	Network Administration, Service Oriented Architecture	FTF
EXPERT Training		
GIS III	Provision of WMS, WFS, WCS and other needed services	FTF & OJT
	Editing and Publishing of Metadata	FTF & OJT
Schema Mapping	Introduction to domain models : INSPIRE, those maintained at EGDI level (from EU projects), GeoSciML, Earth-Resource ML, …	S
	Schema Mapping for Technicians	FTF
	Schema Mapping for Geoscientists	FTF
Operational Management	Project Management for GIS	FTF
Data Quality II	Data quality assurance	FTF
User Access and Licensing	Management of user and access rights	FTF
	Management of pricing and licensing options	FTF
IT-Network II	Maintenance of multi-tier architectures	FTF
	Building an EGDI node	FTF

Table 2: List of training modules and topics and its proposed methodologies for AEGOS adapted to EGDI context

(OJT=on-the-job training, FTF=face-to-face training, S=seminar, ID=info day)