FP7-INFRASTRUCTURES-2012-1

Grant Agreement no. 312845

Scoping Study for a pan-European Geological Data Infrastructure

D 4.1

Report on evaluation of existing interoperable infrastructures

Deliverable number	*D4.1*
Dissemination level	*Restricted*
Delivery date	*30 December 2012*
Status	*Final*
Author	*Jean-Jacques Serrano, BRGM*

SEVENTH FRAMEWORK PROGRAMME

EGDI-Scope Project, WP4

Report on evaluation of existing interoperable infrastructures

Table of Contents

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 fulfil 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, and it will build 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.

The first task, "Evaluation of existing interoperable infrastructures", reviews the status and practices of the technical architecture and design of large initiatives dealing with geospatial information (INSPIRE, GEOSS, EuroGEOSS, GMES,US-GIN, EPOS). It will also evaluate experiences about design, implementation and operation gained through previous EU funded projects including OneGeology-Europe, EuroGeoSource, eWater, EModnet, GeoMind and others. Finally, this task will explore possible synergies with other activities in the e-infrastructures domain. Within this framework, this report provides a broad inventory of relevant existing infrastructures and on-going initiatives, preparing a basis for further analysis of technical requirements of an EGDI.

2 Methodology

Many projects have been selected when they deal with geological information, but also projects from other thematic domains if their architecture could be of interest for EGDI, as EuroGEOSS for example, which addresses biodiversity, forest, and drought domains. This architecture seems very similar to the one we could expect from EGDI.
For many projects, the architecture is not the main goal, as they have more scientific objectives. Also the technical aspects are not always well described.

Other very interesting sources of information to design an architecture are the international initiatives (as OneGeology, GEOSS, EPOS, …). As they have to work with a lot of data/services providers they must specify and describe the rules for the partners to collaborate to the infrastructure. Then technical documents are available describing rules, components, but also recommendations and implementation plans.

A third category of information is provided by the directives (as INSPIRE) and framework documents related to European activities (as GMES or SEIS).

The last category comprises the standards developed by ISO, OGC, CEN in the spatial domain. They described specifications of components (mainly services but not only) and also methodology or plan to design an SDI.

The list of projects, initiatives, directives and standards is presented in the table of the next section, and a summary of each project is presented in the annex, with a comment.

From the analysis of these existing projects, initiatives, and documents a list of 16 topics or components that have to be considered when designing the EGDI architecture in phase 2 has been made. These topics are listed below, a description of each topic or component is provided in section 4:

- Principles
- SDI Phases
- Project documents
- Thematic domain policies
- Types of infrastructure
- Conceptual Frameworks
- SDI Reference Model Components
- SDI Reference Model Services
- Interoperability / Harmonisation
- Multilingualism
- Methodology to develop a SDI
- Technology
- Connection with thematic tools
- Web 2.0
- Sustainability
- EGDI Portal

3 Projects, initiatives, directives reviewed

A summary of each project and a comment are provided in Annex A.

Projects
AdaptAlp
AEGOS
BLAST – Bringing Land and Sea Together
COOPEUS
DRIHMS
EarthQuake Data Portal

EGIDA
EMODNet
EMSO- European Multidisciplinary Seafloor Observatory
ENVRI- Implementation for a cluster of ESFRI Infrastructures
EOMINERS
ESDIN (EuroGeoGraphics)
EUCoRes – European Coal resources
EuroGeoSource
EuroGEOSS
EWater
GBIF
GENESI-DEC – A single access point to Earth Science data
GeoMind
GeoMol
Geo-Seas
GEOWOW
GS Soil
Helix Nebula
iCordi
InGeoCloudS - INspired GEOdata CLOUD Services
LESSLOSS
NERA – Earthquake risk assessment and mitigation
Ocean Data Interoperability Platform (ODIP)
OneGeology-Europe
OpenQuake / GEM (Global Earthquake model)
PanGeo
Planet Data
ProMine
SAFER – Seismic early warning for Europe
SafeLand
SciDIP-ES
SHARE – Seismic Hazard Harmonization in Europe
SUBCOAST
Terrafirma
ThermoMap
VERCE
WMO Information System
Initiatives
EarthCube
EPOS – European Plate Observing System
EUDAT – European Data Infrastructure
Eye on Earth
GEOSS

Google
US- NGDC - National Geophysical Data Center
OneGeology
ORFEUS - Observatories for European Seismology
US-GIN: Geoscience Information Network
USGS – SDI
Directives, framework documents
GMES
European Interoperability Framework (EIF)
INSPIRE
SEIS - Shared Environmental Information System
Standards
CEN- Geographic Information - Spatial Data Infrastructures
OGC/ISO – Geographic Information
ISO - RM-ODP: Reference Model

4 Infrastructure topics and components to consider

This chapter describes the components and topics to consider when specifying an infrastructure (during the second WP4 task “Technical specifications”). All of them come from the analysis of projects, policy documents, and standards during the first WP4 task “Evaluation of existing interoperable infrastructures”.

Several SDI definitions exist, but we could adopt these ones for EGDI:

SDI Definition (Doug Nebert, 2004):
“A spatial data infrastructure (SDI) is a framework of spatial data, metadata, tools, and a user community that are interactively connected so that spatial data can be used in an efficient and flexible way.”

USGS: An essential function of an SDI is to provide an overall framework and architecture within which new applications can be developed and integrated.

Research Infrastructures have a broader scope than SDIs:

e-Infrastructure
“e-Infrastructure refers to a combination and interworking of digitally-based technology (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organizational structures needed to support modern, internationally leading collaborative research be it in the arts and humanities or the sciences.”
(From ENVRI project – http://envri.eu)

4.1 Principles

Before designing an infrastructure, some main principles are defined, as for example those for INSPIRE (or SEIS with almost the same principles), GEOSS-Infrastructure, ….

INSPIRE Common principles:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.

- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

INSPIRE-compliance is one of the main requirements for an EGDI. Building on the experience of existing and ongoing projects and initiatives the consequences for an EGDI have to be elaborated in the next stage of EGDI-Scope, connecting technical requirements with functional and legal requirements.

EIF (European Interoperability Framework) 12 principles for public administrations

Main principle of GENESI-DEC:
The infrastructure with his future extensions will be able to deliver reliable and effective services supporting the diverse research communities providing at the same time innovations that allow usage in an easy and interoperable way.

GEOSS – Infrastructure:

- Architecture: Achieve sustained operation, continuity and interoperability of existing and new systems that provide essential environmental observations and information, including the GEOSS Common Infrastructure (GCI) that facilitates access to, and use of, these observations and information.
- Data management: Provide a shared, easily accessible, timely, sustained stream of comprehensive data of documented quality, as well as metadata and information products, for informed decision-making.

4.2 SDI Phases

The development of most of infrastructures have been split into several phases. The general phases are the following:

- Concept and design, including analyses, methodologies, reference models, state of play related to main SDI areas
- Implementation, including development methodologies, development implementation, deployment, management and maintenance
- Validation, including monitoring, testing and reference standards compliance

4.3 Project documents

The documents delivered by the projects often focused on the objectives (scientific delivery – data, methodologies, models, results) but they don't describe the architecture in details when it is not a key component of the project.

The most interesting documents for EGDI Architecture are the technical documents provided by international initiatives, or policy documents as the rules for collaborating have to be described to a large community of data and services providers (example of INSPIRE, GEOSS, OneGeology, USGS-SDI, …).

Some of them give lessons learned during the implementation that should be considered by EGDI (example of SEIS).
The third category of documents are the standard specifications provided by ISO, OGC, CEN which describe the technical interfaces of services, the metadata and data structure, but also a reference model for SDI (CEN/TR 15449).

4.4 Thematic domain policies

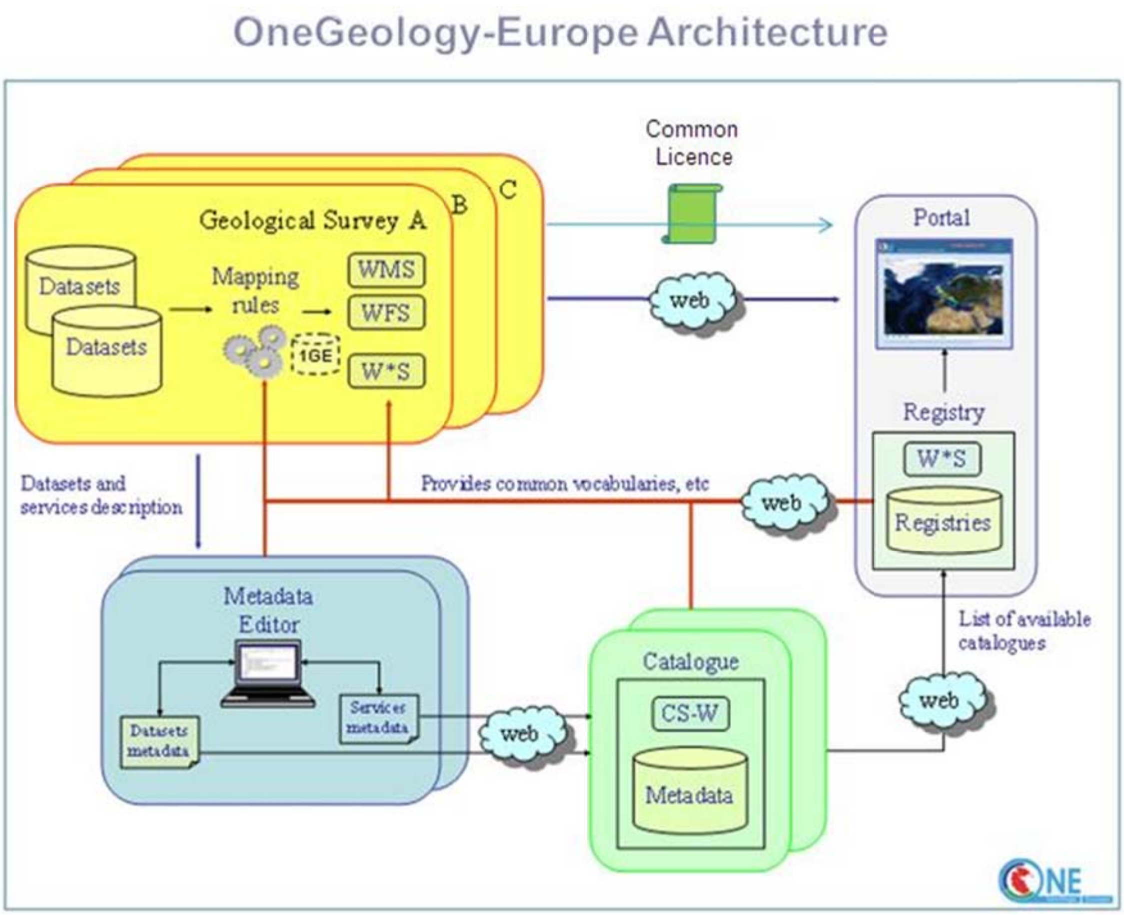
Several European policies exist related to geological domains (groundwater, raw materials, mining waste, …). Early in the technical design phase these policies have to be identified and analysed for their consequences on the architecture (some reporting requirements, mandatory properties, …).

4.5 Types of infrastructure

Some architecture examples are presented to illustrate various possibilities for EGDI.

OneGeology-Europe:

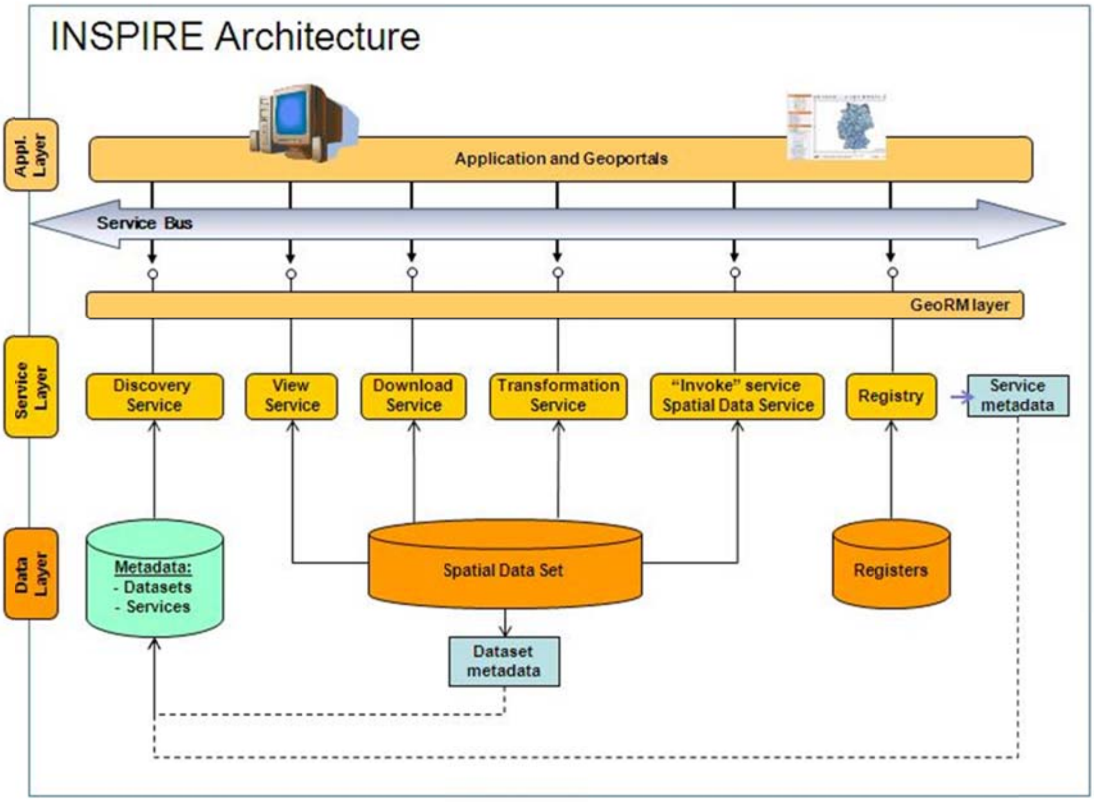
OneGeology-Europe aims to create dynamic digital geological map data for Europe. It makes a significant contribution to the progress of INSPIRE - i.e. develop systems and protocols to better enable the discovery, viewing, downloading and sharing of core European spatial geological data.
OneGeology-Europe addresses licensing and multi-lingual aspects of sharing geological knowledge and demonstrates best practice examples of the delivery and application of geological spatial data in the public and private sectors.



OneGeology-Europe has implemented all technical chapters of INSPIRE: metadata, data, services, data sharing and monitoring. As not all INSPIRE rules were completed at the end of the project (October 2010) the implementation is not 100% INSPIRE compliant but very close to what is/will be required.

INSPIRE:

INSPIRE is an European Directive defining an Infrastructure for Spatial Information in the European Community (http://inspire.jrc.ec.europa.eu/). This directive addresses several technical chapters about metadata, services, and data specifications. For EGDI architecture the main point to consider is the INSPIRE architecture defined by the Network Services team, and presented in the following figure:

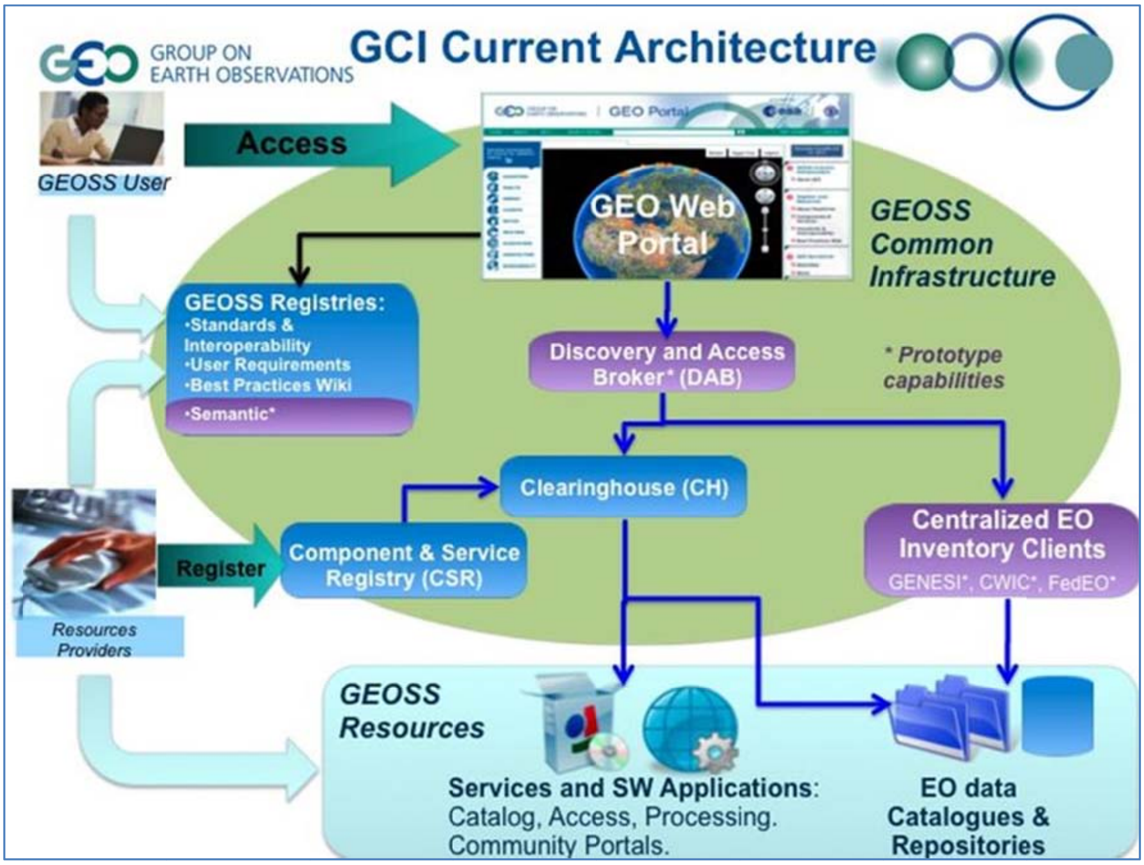


INSPIRE architecture is organised in three layers (the applications, and geoportals layer is not requested by INSPIRE):

- A data layer with all datasets described by their metadata, and the registers
- A service layer with the Network services (discovery, view, download, transformation and invoke service) and the spatial data services. All of them are described by their metadata and process spatial data (or metadata for the discovery service). The registry service is a particular example of spatial data service.
- A GeoRM (Geo Right Management) layer to address when necessary the issues of authentication, payment, …

For EGDI, the INSPIRE compliance is a strong requirement.

GEOSS: GCI (GEOSS Common Infrastructure)



The GEO Web portal provides access to many resources thanks to a flexible connector the Broker (composed by several brokers to discover, access, process data). Resources are data, services, registries, portals, …

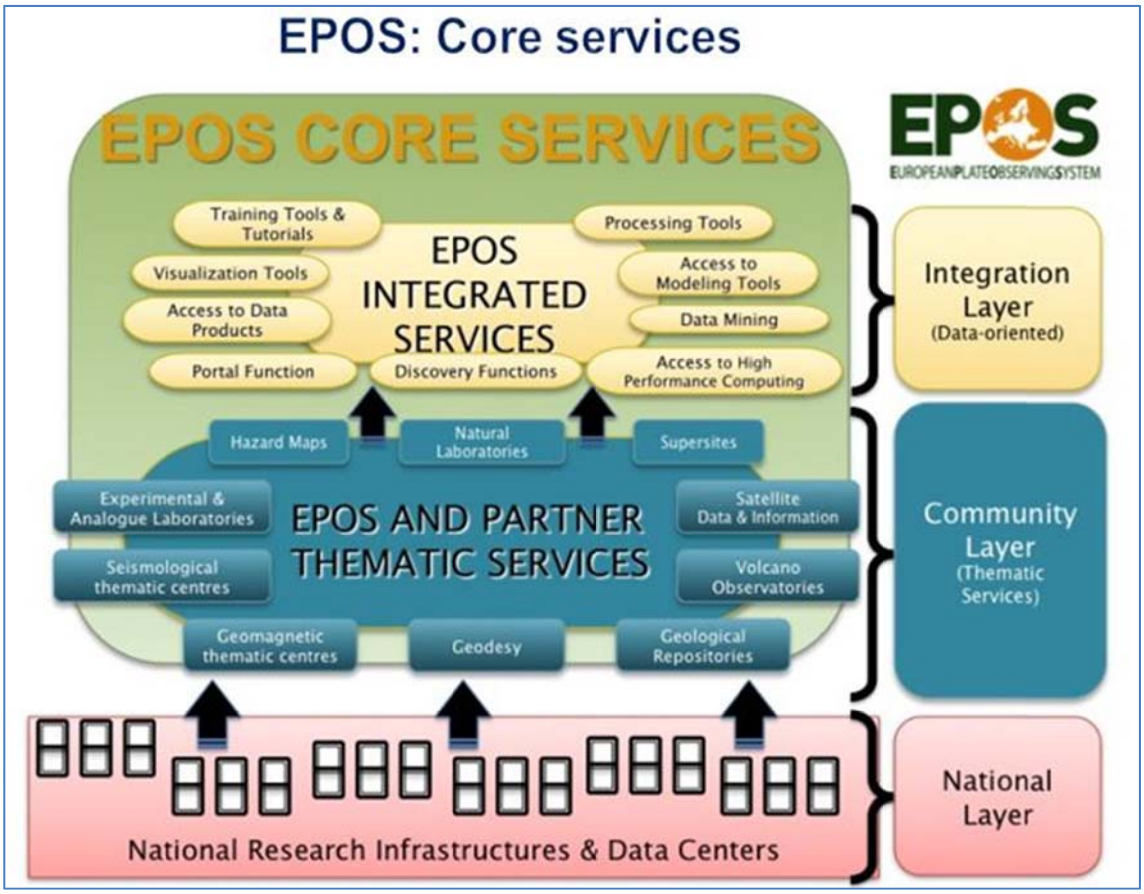
From the GEO Web portal a user can discover, access and use GEOSS Resources such services and software providing access to catalogues, processing services, portals; and Earth Observation data catalogues and repositories.
To make this possible, some common rules and components have to be implemented:

- A clearinghouse where providers register resources (thanks to registry services). Resources are services, software
- Another way to access resources is to use EO clients, also registered in a centralized inventory
- As these two ways to access resources use different standards and formats, a brokering approach has been implemented with a Discovery and Access Broker enabling the portal to be independent from these various “catalogues”.

The portal gives also access to registries about common rules, standards that have to be used to improve access and interoperability.

EPOS:

The European Plate Observing System (EPOS) is the integrated solid Earth Sciences **research infrastructure** to promote and make possible innovative approaches for a better understanding of the physical processes controlling earthquakes, volcanic eruptions, unrest episodes and tsunamis as well as those driving tectonics and Earth surface dynamics



The EPOS Core Services will provide the top-level service to users including access to multidisciplinary data and metadata, virtual data from modelling and solid Earth simulations, data processing and visualization tools as well as access to high-performance computational facilities. EPOS will enhance data processing and modelling capacity and capability as well as develop new theoretical and numerical tools to harness computational power in a distributed European architecture.

To give access to resources provided by national research infrastructures and data centers, EPOS has defined two layers of services:

- The thematic services are to make thematic data available to EPOS from a specific community, and provide specific services to a specific community. These services also include cross-disciplinary integrated observatories providing applications that integrate data and deliver products.
- The integrated services are IT services and services providing access to interdisciplinary data and information from all the EPOS thematic services in an integrated fashion.

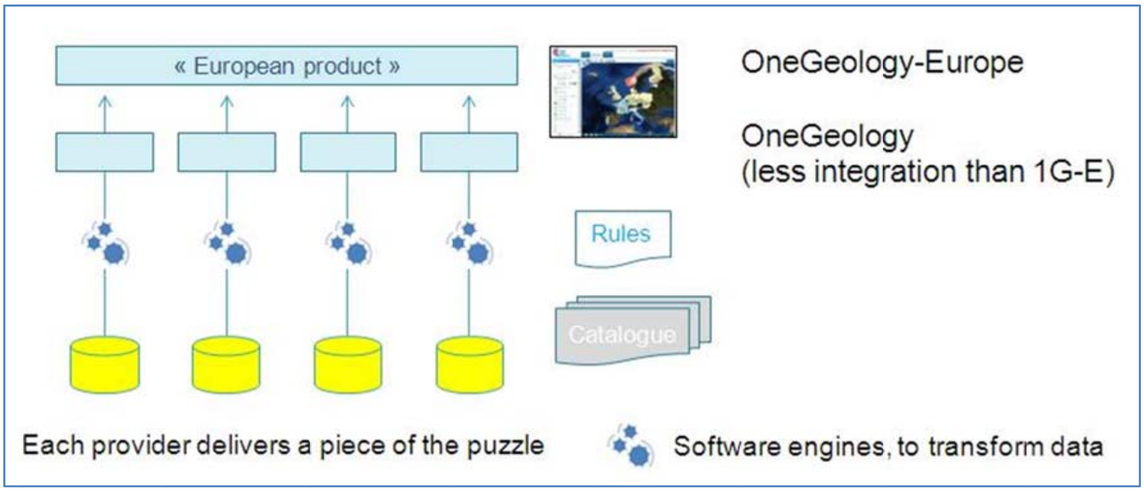
They consist of discovering, viewing, accessing, processing data, and also accessing to HPC.

Possible types of infrastructures

From these previous examples (and more, according to all projects reviewed), and completed by ProMine and EuroGeoSource architectures, we can think for EGDI architecture to some schemes:

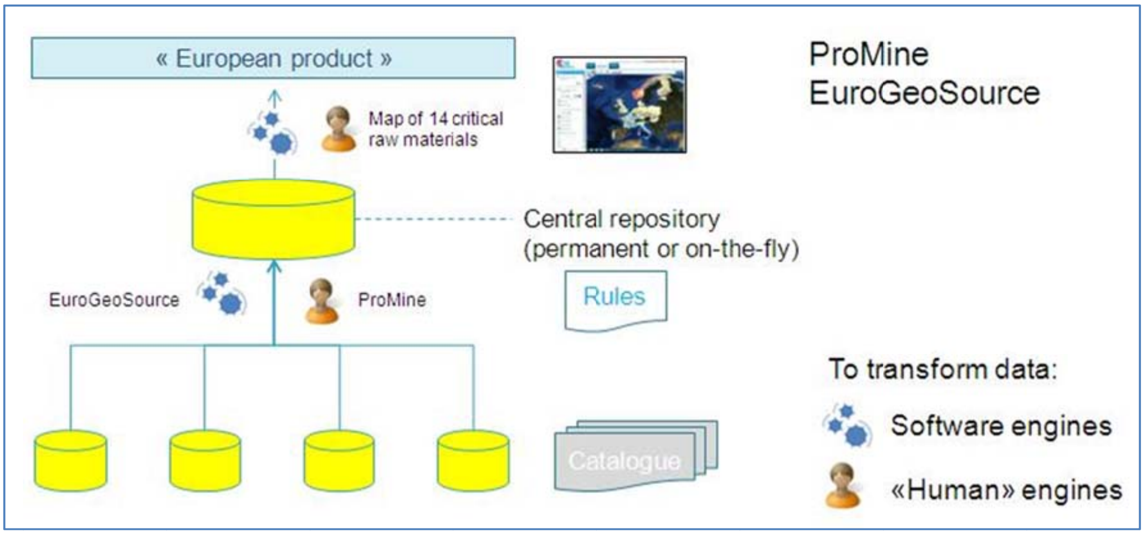
- Architecture Type OneGeology-Europe:
- Architecture Type ProMine, EuroGeoSource:
- Architecture Type EuroGEOSS, GEOWOW, …
- Architecture Type EPOS (Research Infrastructure)

1. Architecture Type OneGeology-Europe:



In this architecture, each provider delivers data and services as a part of the global puzzle according to common rules. The final product is made by a “simple” application or portal requesting the providers (which services are registered in a catalogue) to deliver its part. Engines are necessary to transform local/national data into the common data model (including semantic aspect). INSPIRE-compliance was one of the main requirements for OneGeology-Europe, which has been partly achieved. For example an important issue still to be solved is the continuation and governance of the infrastructure.

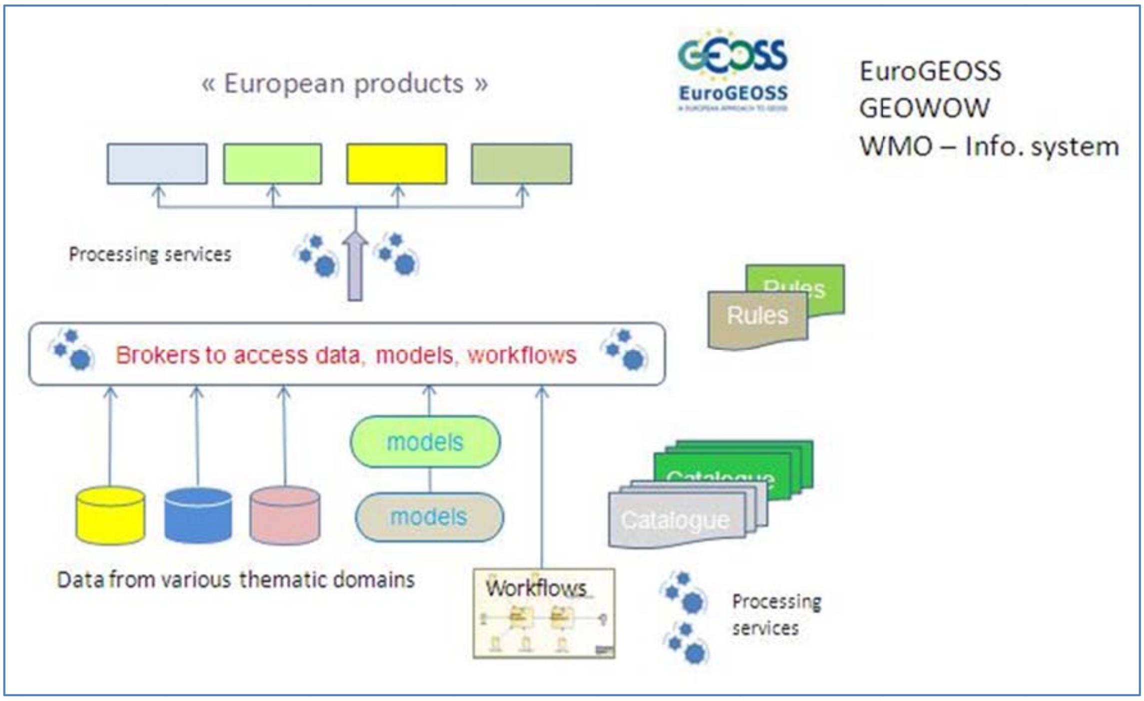
2. Architecture Type ProMine, EuroGeoSource:



In this architecture, each data provider delivers the data to a central system (by automatic harvesting – EuroGeoSource - or manual importation – ProMine). The INSPIRE requirements (for interoperability) are taken into account by the central system which offers view and download services to make an European product (as the map of 14 critical raw materials).

From the perspective of user performance and requirements this architecture has reached a high level of INSPIRE compliance, but legal issues concerning data delivery from some involved Member States, as well as the governance and continuation (like OneGeologyEurope) are issues to be solved yet.

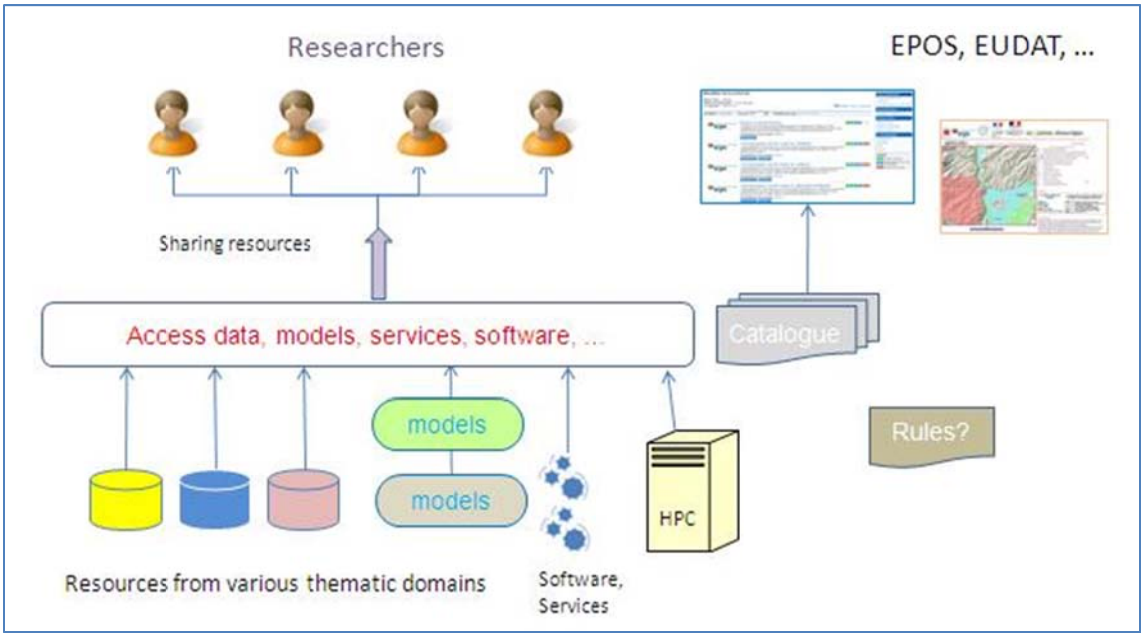
3. Architecture Type EuroGEOSS, GEOWOW, …



This architecture provides more components and is more flexible than the others presented. The broker approach allows to access more resources as the interoperability constraints are managed by the brokers. It offers also access to models to process data, according to workflows described and registered.

Several mechanisms and components (semantics discovery, harmonisation processes, use of coverages, …) makes interoperability between different domains possible for combining data to deliver a result.

4. Architecture Type EPOS (Research Infrastructure)



This type of infrastructure (research infrastructures) represented by EPOS seems to be different from EGDI perspective. The main goal of RI is to help researchers to share resources (data, services, models, software, computers, …) for their work instead of EGDI which is more to deliver a geological product well specified by processing data according to shared rules.
However, EGDI will have to communicate with RI to offer geological layers as reference layers, and possibly to use some resources provided by RI (EPOS in particular). This relationship will have to be specified during the EGDI technical design phase.

Another classification of infrastructures are also suggested:
- Data and information infrastructures: infrastructures to deliver data (with services) as most of the projects analysed have implemented
- Model and simulation infrastructures: infrastructures to use data and process them with models
- Planning and decision support infrastructures: infrastructure to use models to provide information for decision.

USGS defines two types of SDI:

SDI as an Applications platform
An essential function of an SDI is to provide an overall framework and architecture within which new applications can be developed and integrated
Using the SDI as a community applications platform would allow users to take advantage of existing applications that perform functions that they need rather than having to develop

their own applications. In turn, when users develop new or improved services, they could more easily make them available to others through the SDI.

SDI as a Workflow platform

An important benefit of workflow approaches is the detailed record of data transformation and data-processing that is generated, which is a vital part of tracking the provenance of data and is critical for the long-term curation and reuse of data. Documenting and curating the workflows themselves may be an important role for an SDI in capturing geospatial expertise and ensuring the long-term reusability of the spatial data supported by the SDI.

4.6 Conceptual Frameworks

Some conceptual frameworks composed of technical documents, standards, rules, tools are available to help the architecture design. The following are suggested, and should be explored during the technical design phase:

- Data collection including sensor networks, crowd sourcing
- Management, preservation, sharing including documenting resources with metadata, preserving information for long-term but also for traceability, linking open data to share resources, core services to support discovery, view and access to resources
- Analysis and modeling making data from different communities comparable before combining them in forecasting and simulation models
- Decisions and actions focusing on the value added service chains
- Communication including new ways of communication in terms of devices but also of processes
- Security: for single sign-on, access control management, guidelines for configuration, deployment, …

4.7 SDI Reference Model Components

Many well-known components are implemented in most of infrastructures:

- Data specifications: to describe data structure, related metadata, data quality, encoding, portrayal rules
- Registers: to describe and publish resources
- Discovery: to search for and discover resources
- View: to view resources
- Download: to access and exchange resources
- Invoke: to interact with resources
- Orchestration – Composition: to provide aggregated resources including workflows for services composition
- Rights management: to manage access rights to resources
- Links data/services to (legal) documents
- Ordering and payment components
- Data capture from distributed sensors (real time monitoring)

4.8 SDI Reference Model Services

To classify and organise services in the infrastructure, the ISO standard for services (ISO19119) suggests this services taxonomy:

- Human interaction services, as catalogue viewer, view services
- Model management services, as catalogue service, registry service, feature or coverage access services, gazetteer service
- Workflow / task services, as chain definition service, workflow enactment service
- System management services, as authorization, authentication services, performance measurements
- Processing services, related to spatial, temporal and thematic areas as coordinates transformation service, feature generalisation service
- Communication services, as encoding service, format conversion service

Another services classification provided by USGS:

- Data transformation: to convert data between different formats, encoding systems, coordinate systems, or between different levels (aggregation or disaggregation); gridding algorithms to convert vector-based to raster-based data formats; semantic translation services
- Data integration: services built on data transformation services to support assembly of data from different sources into a combined dataset
- Spatial and statistical analysis services
- Modeling services
- Visualisation services

4.9 Interoperability / Harmonisation

Interoperability levels:

The practical implementation of the conceptual model for cross-border/cross-sectorial services requires each of these levels to be taken into account:

- Legal interoperability to align legislation so that exchanged data is accorded proper legal weight,
- Organisational interoperability to coordinate processes in which different organisations achieve a previously agreed and mutually beneficial goal,
- Semantic interoperability to precise meaning of exchanged information which is preserved and understood by all parties,
- Technical interoperability to plan technical issues involved in linking computer systems and services

Many projects have addressed the technical and partly the semantic levels, using standards for metadata, data models, services and common vocabularies/ontologies with related services for the semantics.
The organisational and legal levels are not really implemented.

Multidisciplinary interoperability:
Interoperability should be also implemented between various thematic domains, to make possible by the users the combination of different data sources. From a technical point of view, a good way to address this is to add semantics capabilities to search and discovery services, and to use coverages (gridded data), already experimented in several projects (as EuroGEOSS to combine forest and biodiversity data).

Interoperability with external e-infrastructures
As described in the infrastructure type section, EGDI and Research Infrastructure have partly different objectives, but they have to establish links to collaborate, especially in the future to address issues which are not known nowadays. One of the main domain-related RI to EGDI is EPOS. The first relationship could be for EGDI to deliver reference European geological product (as services) to be used by EPOS researchers. In the other direction, EGDI could access EPOS observatories data.
From a technical point of view that means to agree on rules related to the services interface for a seamless integration into both infrastructures.

4.10 Multilingualism

In the European context, addressing multilingual issues is a strong requirement. Multilingualism must be taken into account in various components of the architecture (processes, metadata, data and user interfaces, …):

- Portal / Applications (user interface), to make portals and applications available for various languages
- Data, to manage texts in various languages where practicable
- Code-lists / Vocabularies, to provide code-lists in various languages, and related services to use them
- Legends (maps, boreholes, 3D models, …), to provide legends in various languages

Technical solutions have been implemented at least in OneGeology-Europe.

4.11 Methodology to develop a SDI

To develop a SDI we can refer to some international standards and to implementation plans defined by some organisations (GEOSS, USGS, …)

Reference Model for Open Distributed Processing

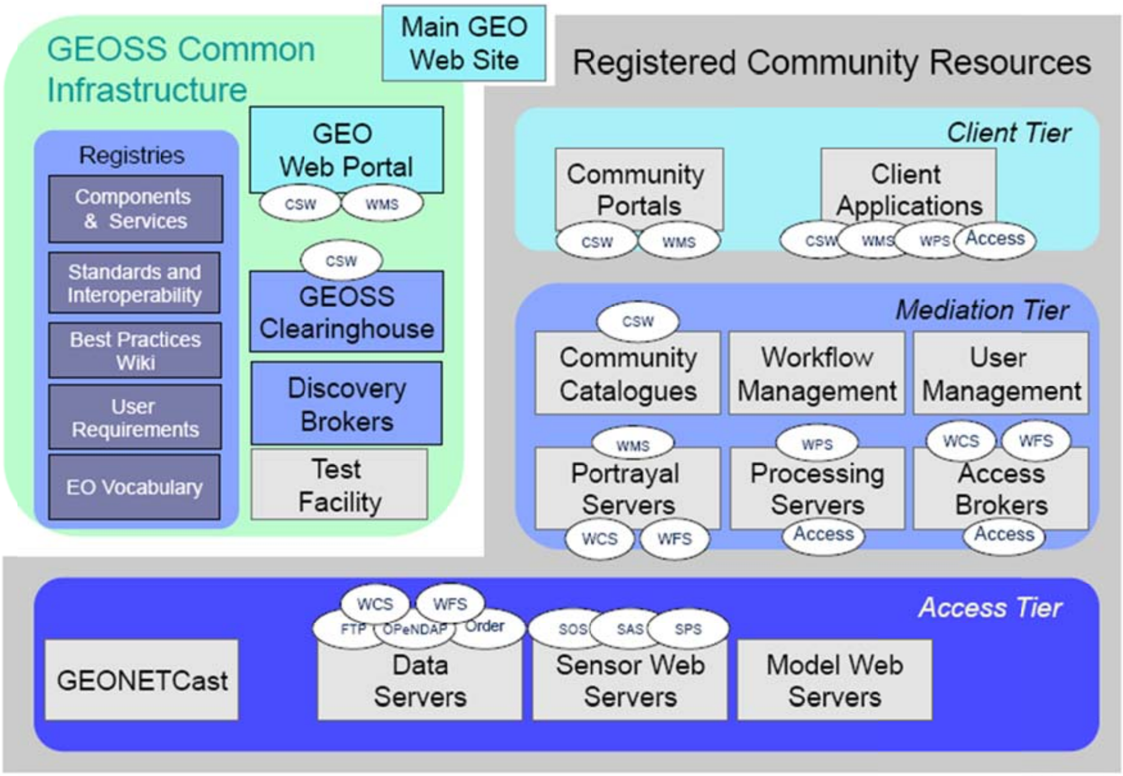
This model is described in the ISO RM-ODP standard (Reference Model for Open Distributed Processing) with its 5 viewpoints: Enterprise, Information, Service, Engineering, Technology, completed by the CEN organisation adding an implementation viewpoint in the technical report CEN/TR 15449):

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

GEOSS Architecture implementation plan

The GEOSS Architecture Implementation Pilot – 5 (GEOSS AIP-5) uses the RM-ODP model to specify the content of each viewpoint.

The engineering viewpoint summarized in the figure below is more an organizational scheme than a methodology but could be a good example of framework for EGDI design:



GEOSS AIP-5: Engineering components with services

This GEOSS architecture is organized as a 3-tier model:

- The top tier (client tier) is the only one for users and provides interfaces to describe and use the services offered,
- The middle tier (mediation tier) embodies all the business processes required to respond to requests issued by the clients, from authentication to geoprocessing, from map viewing to statistical charts
- The lower tier (access tier) provides access to data (read and write), and to registries records

All these components have to be registered into the GEOSS registries, and must use standards and interoperability arrangements described in documents also registered. User requirements, best practices and Earth Observation vocabularies are available to all GEOSS users from the registries.

USGS SDI Implementation plan:

Suggested 3 phases by USGS in its implementation plan: preparation and planning; design, development and testing; rollout and refinement.

Preparation and planning

- Review similar SDI
- Determine and define SDI system requirements (on basis of objectives and user needs)
- Determine the organisational structure of the SDI

- Identify goals, establish timeframes and milestones, and develop performance metrics
- Communicate the general outline for implementing the SDI

Design, development and testing
- Identify and define standards to use
- Identify processes and develop software
- When standards, processes and software in place for a SDI prototype, define a training program

Rollout and refinement
- Include a process for rolling out the SDI
- Include a process for fine-tuning the program

SEIS lessons learned:
SEIS (Shared Environmental Information System) described in details ten important aspects to consider: standardisation, collaboration, user friendliness and accessibility, availability of sources, automation of reporting, decentralised integration of services, commitment, financing, managerial issues, technical issues.
All these aspects are described in the annex of this document. For the technical design we should focus on 4 of them: standardisation, automation of reporting, decentralised integration of services, technical issues.

4.12 Technology

From a technological point of view, many projects or initiatives are users of these components:
- Technical architecture: description of components and their organisation. Service Oriented Architecture (SOA).
- Performance (including availability/scalability, …): requirements defining performance criteria and values, and tools to measure and report
- SDI Monitoring: tools to monitor the infrastructure
- Technical frameworks : security framework, metadata framework, …
- Cloud-based computing infrastructures: to host, process, and analyse large volumes of multidisciplinary data, and also facilitate the long-term data preservation
- Semantics technologies: to improve the link between various disciplines
- Database-centric computing: computations brought to the data rather than data brought to the computations
- Data integration on the Web: to determine the best way to aggregate huge amounts of semantically rich information and consider how the resulting information is generated and analysed.
- Use of open source / commercial software
- Devices: devices used in the infrastructure for processing and getting results (desktop, web, mobile)

During the technical design phase, we should check the relevance of these components for EGDI.

4.13 Connection with thematic tools

Even if the general goal is to have services available as web services, for several reasons (performance, access control, …) thematic processes will still be run on local computers to process data. Therefore the EGDI architecture should define how to communicate with the thematic tools as:

- GIS
- Boreholes, wells, sections viewers
- 3D Modelling
- 3D Viewers

That means to define how in the infrastructure the users can get data from these tools or provide data to them, using where possible, standard ways for this communication.

4.14 Web 2.0

Even if the input from citizens (crowd sourcing) is not used in our geological domain, the architecture should have to think about this component, and maybe defined how it should be taken into account for the future.

4.15 Sustainability

Two key points have to be addressed for the EGDI sustainability:

- Governance: organization setup to manage the infrastructure
- Maintenance: how maintenance is organized?

They cover more than the architecture point of view, but we will have to address them to define how the architecture is governed and maintained.

4.16 EGDI Portal

This component will provide an entry point for accessing all EGDI resources:

- Links to resources catalogues,
- Guidelines,
- Best practices,
- Tools,

5 Conclusions

Many projects and some international initiatives exist in the spatial domain, dealing with geosciences. Most of them use the standards specified by ISO/OGC/CEN to improve interoperability within a thematic domain or between domains. The number and the organisation of components used to build their infrastructure could be different, according to their objectives, but we can identify common topics that seem necessary to address and to control to setup an interoperable and open infrastructure.

The analysis identifies that Research Infrastructures (as EPOS) have partly a different objective than EGDI, but the collaboration seems necessary and should be specified.

Very useful framework documents are provided either by global initiatives (GEOSS) or by directives (INSPIRE). Thematic directives or European communications (as the Raw Materials Initiative) will have to be checked for their possible consequences on the EGDI architecture design.

And more than a (classical) technical implementation plan, a well communicated roadmap, and additionally a training program for providers are key elements for the success of the SDI implementation.

Both user requirements from WP2 and all the topics identified in this document will be taken into account for the technical design during the next phase. At this stage of the EGDI-Scope project and from this inventory, the following generic requirements can already be derived, as a starting point for the follow-up analyses:

- INSPIRE compliance, with regard to issues like: metadata architecture and harmonization, data storage, sustainability and governance, data delivery, etc.;
- Thematic focus and classification to be decided, connected to relevant use cases at EU or international level, e.g. concerning geohazards and earth resources (shale gas, minerals)
- Delivery of datasets:
  - Already available datasets at national geological surveys;
  - Datasets and information from relevant European projects and initiatives;
  - freely available datasets, in some cases possibly also restricted datasets;
  - Ranging from raw data to integrated (decision support) models, depending on possibilities and requirements
- Target groups: policy makers at EU-level, industry, research community, general public
- Adapted to actual legal frameworks as far as feasible;
- Governance and funding models to provide continuity

Annex A. Summary of projects reviewed

For many projects, there is not a real infrastructure defined but a portal providing access to the results (maps, data, documents, software, …). However they have been analysed from a technical point of view for the elements they are dealing with, as they have to be taken into account into the EGDI infrastructure.

In the following sections, projects, initiatives, directives are sorted in alphabetic order. For each of them there is a description from their web site, a comment when useful, and the web site URL to be able to get more information when necessary (especially during the design phase).

6 Projects, initiatives, directives

6.1 AdaptAlp

Description:

The main focus was hazard mapping of firstly geological and secondly water-related hazards, as there is a considerable lack of standards and requirements for methods and procedures to develop hazard maps in this field.

The project aimed at providing clear guidance in the field of hazard mapping, improving hazard mapping on a transnational basis, by harmonizing the technical terminology and by creating a basis for cross-sectoral hazard mapping procedures.

A multi-lingual glossary was elaborated to standardize the terminology for geological hazards across the Alpine Space: 97 terms and definitions to geological hazards (languages: English, French, German, Italian, Slovenian and Spanish).

Tools were developed that are able to enhance the mapping of endangered areas and for communication purposes.

http://www.adaptalp.org/

Comment:

Guidance to harmonize hazard mapping is a component of this project. The glossary is also a major component; some questions about it: how to standardize its structure and its access with a web service, to be shared by many users?

There is a need to run models: how are they described (metadata?), used (how are the parameters defined? How are the data provided to the tool?).

6.2 AEGOS

Description:
The AEGOS project aims at setting-up the preparatory phase for the building of an information system containing and making accessible data and knowledge on African geological resources including mineral resources, raw material, groundwater and energy (georesources).

http://www.aegos-project.org

Comment:
The recommendations for the AEGOS architecture are based on the use of INSPIRE and ISO/OGC standards (distributed system). Some existing standards for the data model related to geology, mineral resources, and hydrogeology are also recommended.
The global AEGOS model is compliant with the Reference Model of Open Distributed Processing defined by ISO.

6.3 BLAST – Bringing Land and Sea Together

Description:
BLAST - Bringing Land and Sea Together - was a regional project for better integration of information across the coastal margin in the North Sea region. Over three years, 17 partners from 7 countries, including governmental organisations, universities and private companies, collaborated on the harmonisation and integration of land and sea data.
The Climate Change in the Coastal Zone Work Package developed a web-based decision-support system for Integrated Coastal Zone Management (ICZM) in a climate change perspective.

An Indicator System was also implemented in the decision-support system, permitting users to assess and visualise responses to the impacts of climate change in the coastal zone. Use of the Indicator System was intended to reinforce the role of ICZM in mitigating and adapting to climate change.
The decision-support system was developed based on a conceptual model, an indicator approach and an innovative architecture. The overall aim of the system was to offer specific tools to policy makers and planners, assisting them in making better decisions on mitigation and adaptation measures.

To help improve ICZM-related decision making, BLAST developed a Coastal Indicator System (COINS) for use by planning authorities. Based on the 27 climate change indicators developed by the European Expert Group on ICZM, COINS indicates how coastal planners can take account of potential future impacts of climate change in their work. COINS is web-based, built on open source components and can be used free of charge from most web browsers under different operation systems (http://blastdss.statkart.no/)

http://www.blast-project.eu/

Comment:

- Availability of relevant datasets and feature types
- Identification whether standardized metadata exists for these datasets.
- Setup of the BLAST metadata catalogue, and the collection of metadata.
- Web portal with many layers, possibility to add WMS layers
- Maps of indicators (with selection of values for parameters) (flood hazard cover, flood hazard depth, erosion prediction, …)

6.4 COOPEUS

Description:
The COOPEUS project aims at fostering the EU-USA cooperation in the field of Research Infrastructures dedicate to environmental sciences.
The objectives of COOPEUS can be summarized in the following activities:

- Ensure a larger harmonization and interoperability between research infrastructures across communities and crossing borders
- Exploit synergies between existing initiatives
- Optimize technological implementation by making use of evolving concepts in e-infrastructures

http://www.marum.de/Page12529.html

Comment:
The project is starting (September 2012), then no results are yet available.

6.5 DRIHMS

Description:
Distributed Research Infrastructure for Hydro-Meteorology Study.

The Distributed Research Infrastructure for Hydro-Meteorology (DRIHM) project intends to develop a prototype e-Science environment to facilitate the collaboration between meteorologists, hydrologists, and Earth science experts and provide end-to-end HMR services (models, datasets and post-processing tools) at the European level, with the ability to expand to global scale.
The objectives of DRIHM are to lead the definition of a common long-term strategy, to foster the development of new models and observational archives for the study of severe hydro-meteorological events, to promote the execution and analysis of high-end simulations, and to support the dissemination of predictive models as decision analysis tools.

http://www.drihms.eu/

Comment:
Prediction of floods and other hydro-meteorological events relies on hydrological and meteorological forecast models. These predictions are based on observational measurements, There is a need to understand the entire forecasting chain, from observations through to civil defence response, resulting in complex workflows able to combine different data sets, models and expertise in a flexible manner
From a technical view, the project is focusing in Grid and High Performance Computing (HPC).

6.6 EarthQuake Data Portal

Description:
The Earthquake Data Portal is the rendering layer of an integrated Infrastructure that enables the research community to have access to a broad range of earthquake data from Europe and its surroundings.

http://www.seismicportal.eu

Comment:
Portal providing access to many sources of data / information, archives (as the European archive of historical earthquake data – AHEAD)
Portal developed in the NERIES project and provides access to several databases
The Portal provides a single point of access to diverse, distributed European earthquake data provided in a unique joint initiative by observatories and research institutes in and around Europe. Based on internet-standard portlet and web services technologies, it enables the scientists/users to integrate and combine different data services.
Provides access to web services about events, historical earthquake data, waveform, and client tools to retrieve data.
Some web services are OGC compliant – WMS and WFS – based on the GeoServer framework (http://www.seismicportal.eu/jetspeed/portal/explorers/ogc.psml). Event information uses the QuakeML language to deliver data.

Provides also a link to the European Seismological Reference Model (ESRM). Seismological models of Earth crust and upper mantle structure are critical for many basic and advanced seismological tasks, such as the calculation of earthquake source parameters (location, magnitude, geometry), shake-maps, and the imaging of lithospheric dynamic processes.

6.7 EGIDA

Description:
EGIDA will deliver evaluation processes, tests and assessment indexes, expertise databases, a "geo label" concept, surveys and other instruments designed to help link relevant European S&T communities to GEOSS and ensure it is built using state-of-the-art science and technology
EGIDA ("Coordinating Earth and Environmental cross-disciplinary projects to promote GEOSS") has defined a general methodology for a sustainable contribution to GEOSS (Global Earth Observation System of Systems). The EGIDA Methodology is a general methodological approach for implementing a (re-)engineering process of the S&T national infrastructures and systems, which can be adopted by national/regional S&T communities, for a sustainable contribution to the GEOSS and relevant European initiatives based on a SoS approach, through the mobilization of resources made available from the participation in national, European and international initiatives and projects.

http://www.egida-project.eu/

Comment:
EGIDA is more a methodology and networking to support GEOSS than an infrastructure, but it will be of interest for the re-engineering processes.

6.8 EMODNet

Description:
European Marine Observation and Data Network
The EMODNET-Geology project is one of six preparatory action projects that, in addition to marine geology, bring together information on marine chemistry, marine biology, hydrography, sea-bed habitats and physical properties. Each project defines the processes, technologies and approximate costs of implementing a fully functioning European Marine Observation and Data Network. For the EMODNET-Geology project, the project partners are compiling data layers for the Baltic Sea, Greater North Sea and Celtic Sea.

http://www.emodnet-geology.eu/

Comment:
The data are visualised in the OneGeology-Europe portal, provided as WMS layer of EMODNet-Geology.

6.9 EMSO- European Multidisciplinary Seafloor Observatory

Description:
EMSO is a large-scale European Research Infrastructure in the field of environmental sciences. EMSO will be based on a European-scale network of seafloor observatories and platforms with the basic scientific objective of long-term monitoring, mainly in real-time, of environmental processes related to the interaction between the geosphere, biosphere, and hydrosphere, including natural hazards. It will be a geographically distributed infrastructure composed of several deep-seafloor observatories, which will be deployed on specific sites around European waters, reaching from the Arctic to the Black Sea passing through the Mediterranean Sea, thus forming a widely distributed pan-European infrastructure.
The vision of EMSO is to allow scientists all over the world to access observatories data following an open access model. EMSO will deliver multi-parametric, long-term (years) time series addressing the seabed and the water column.
EMSO observatories will be equipped with a common set of sensors for basic measurements and further sensors for specific purposes defined by the users.

http://www.emso-eu.org/management/

Comment:
EMSO will implement a Research Infrastructure, some data themes could be useful for EGDI, especially data from observatories. The search engine should be also of interest.

6.10 ENVRI- Implementation for a cluster of ESFRI Infrastructures

Description:
Implementation of common solutions for a cluster of ESFRI infrastructures in the field of "Environmental Sciences" (ENVRI)
The central goal of the ENVRI project is to implement harmonised solutions and draw up guidelines for the common needs of the environmental ESFRI projects, with a special focus on issues as architectures, metadata frameworks, data discovery in scattered repositories, visualisation and data curation.
The collaborative effort has to ensure that each infrastructure can fully benefit from the integrated new ICT capabilities beyond the project duration by adopting the ENVRI solutions as part of their ESFRI implementation plans.

Two use cases, led by the most mature infrastructures, will focus the development work on separate requirements and solutions for data pre-processing of primary data and post-processing toward publishing. The project will be based on a common reference model created by capturing the semantic resources of each ESFRI-ENV infrastructure. This model

and the development driven by the testbed deployments result in ready-to-use systems which can be integrated into the environmental research infrastructures.
The project puts emphasis on synergy between advanced developments, not only among the infrastructure facilities, but also with ICT providers and related e-science initiatives. These links will facilitate system deployment and the training of future researchers, and ensure that the inter-disciplinary capabilities established here remain sustainable beyond the lifetime of the project.

http://envri.eu/

Comment:
The focus of ENVRI on issues as architectures, metadata frameworks, data discovery, visualisation and data curation to implement harmonized solutions is of high interest for EGDI, and its relationship with Research Infrastructures.

6.11 EOMINERS

Description:
Earth Observation for Monitoring and Observing Environmental and Societal Impacts of Mineral Resources Exploration and Exploitation.
EO-MINERS will use current knowledge and data, along with existing and new technological and scientific earth observation-based methods and tools, to monitor mineral resources exploration and mining from concept to closure and observe, monitor and provide information to manage its impacts on the environment and society.

http://www.eo-miners.eu/

Comment:
The interest for EGDI is to check if it will be possible to manage the kind of data (DEM or airborne imagery) and models (to use for providing derived products), and indicators for environmental impacts.

6.12 ESDIN (EuroGeoGraphics)

Description:
ESDIN: European Spatial Data Infrastructure
ESDIN is a bridge between the theory and the practice — from the INSPIRE Directive towards implementation and usage of interoperable geographical data by Spatially enabled Societies, and a challenge to achieve the target — to enable the access to the geospatially referenced data for a INSPIRE Annex1 themes at a European or local level.
Through organisational development and better implementation of existing technology it would be possible to improve the interoperability of, and cost efficiency with which, the existing datasets are maintained and to increasingly improve access to other reference themes at different levels of detail.

Comment:
ESDIN has implemented an architecture compliant with the INSPIRE requirements. It addresses also some complementary topics as a framework for pricing and licencing policy; quality model guidelines; authentication.
The architecture has been developed based on the RM-ODP approach, and has shared the EIF (European Interoperability Framework) objectives.

6.13 EUCoRes – European Coal resources

Description:
EuCoRes project aims for the creation of a geographical database and map of EU coal basins including potential sources of coal bed methane based on a harmonised typology.

The focus of this project is a thorough classification and mapping of coal and coal bed methane (CBM) in the EU. Next to the formulation of a unified classification and terminology, an extensive geographical database is to be constructed where detailed and georeferenced information of all existing coal deposits in the EU is included.

http://ec.europa.eu/energy/coal/eucores/eucores_en.htm

Comment:
EUCoRes provides one database for data related to coal basins, and an harmonised terminology and classification.

6.14 EuroGeoSource

Description:
EuroGeoSource is a data portal, which allows access by Internet to the aggregated geographical information on geo-energy (oil, gas, coal etc.) and mineral resources (metallic and non-metallic minerals, industrial minerals and construction materials: gravel, sand, ornamental stone etc.), coming from a wide range of sources in a significant coverage area of Europe (ten countries).
The EuroGeoSource system will implement content-specific and user-oriented GIS map services on the Internet , based on an inventory and analysis of geo-energy and mineral resource data sets existent in the project countries, together with the user needs regarding these data
The portal will be based on open-source software and standards and will include also an ISO-compliant metadata catalogue.
The following elements characterize the EuroGeoSource portal:
- Multilingual interface of the map viewer (English plus the languages of the project countries)

- Possibility of displaying the data on various backgrounds, using the already implemented map services developed in previous eContent Plus Program projects (eEarth, eWater, Geomind, EuroGeoNames, OneGeologyEurope)
- Harmonization of geo-energy and mineral resource data coming from various countries based on key economic and geological parameters
- Multilingual legend of the data sets using geological dictionaries built especially for the project.

Comment:
Architecture very similar to the one of OneGeology-Europe, it is a distributed system to get data but with a central database (automatically populated by harvesting processes) used to provide data or maps for the web services.
From a technological point of view EuroGeoSource uses a cloud system.

6.15 EuroGEOSS

Description:
The EuroGEOSS project builds an initial operating capacity for a European Environment Earth Observation System in the three strategic areas of Drought, Forestry and Biodiversity. It then undertakes the research necessary to develop this further into an advanced operating capacity that provides access not just to data but also to analytical models made understandable and useable by scientists from different disciplinary domains.
This concept of inter-disciplinary interoperability requires research in advanced modelling from multi-scale heterogeneous data sources, expressing models as workflows of geo-processing components reusable by other communities, and ability to use natural language to interface with the models.

http://www.eurogeoss.eu/

Comment:
EuroGEOSS addresses 3 thematic domains (Forest, Biodiversity and Drought) and provides an architecture and tools to make the data combination possible, as well as the use of models to process these data.
A key component to discover, access, and process data is the EuroGEOSS **broker**, a component (or a group of components) offering many interfaces to reduce the interoperability burden on data providers and applications. In a brokering infrastructure several functionalities can be implemented in the mediation part (discovery, access but also service composition, semantic finding, quality-based ranking, …).

6.16 EWater

Description:
The main objective of the project "Multilingual cross-border access to ground water databases" (eWater) is to increase the cross-border availability, accessibility and re-usability of spatial data on quality, location and use of subsurface waters. The objective of the project will be achieved by developing a WEB GIS portal for hydrogeological data of the participating countries. The envisaged cross-border portal is meant for EC itself, national and river basin water authorities, water suppliers, added-value data service providers, insurance companies, planning and controlling organizations, as well as the general public
Since eWater architecture complies with INSPIRE policy, new data suppliers will be able to distribute their data via the system in future. Moreover, the spatial hydrogeological data infrastructure developed by eWater project, can be used by other international applications, including Water Information System for Europe (WISE).

http://ewater.geolba.ac.at/

Comment:
The project architecture is compliant with INSPIRE architecture: a distributed system with catalogues of metadata (harvested into a central catalogue), services to view hydrogeological maps, and access services to deliver hydrogeological point data (even if the data delivery service is not OGC compliant).
The system is multilingual for metadata, maps, and application, also for data when using code-lists translated (for metadata, and point data)

6.17 GBIF

Description:
The Global Biodiversity Information Facility (GBIF) was established by governments in 2001 to encourage free and open access to biodiversity data, via the Internet. Through a global network of countries and organizations, GBIF promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet.
As a mega-science initiative, GBIF aims to provide an essential global informatics infrastructure for biodiversity research and applications worldwide.

http://www.gbif.org/

Comment:
The infrastructure is divided into six major components: publishing, discovering, indexing, integrating (with other systems), retrieving, analysing

A key focus of the development of GBIF infrastructure has been on the development of a complete catalogue of organism names

6.18 GENESI-DEC – A single access point to Earth Science data

Description:

The project will establish open data and services access, allowing European and worldwide Digital Earth Communities to seamlessly access, produce and share data, information, products and knowledge. This will create a multi-dimensional, multi-temporal, and multi-layer information facility of huge value in addressing global challenges such as biodiversity, climate change, pollution and economic development. GENESI-DEC evolves and enlarges the platform developed by the predecessor GENESI-DR project by federating to and interoperating with existing infrastructures.
GENESI-DEC involves key partners of ESFRI projects and collaborates with key actors of Digital Earth and Earth Science initiatives, including the International Society of Digital Earth and GEO-GEOSS. Thus efficient use of already existing and planned developments is guaranteed.
The objectives of GENESI-DEC are:

- Enlarge Infrastructure: To enlarge the existing GENESI-DR infrastructures in terms of data, resources availability and geographical extent.
- Guarantee Service: To provide guaranteed, reliable, easy, effective access to a variety of data, facilities, tools and services to an ever increasing number of Digital Earth users from all disciplines.
- Harmonise Federation: To harmonise operations at selected key Digital Earth infrastructures limiting fragmentation of solutions.
- Enable User Collaboration: To enable multidisciplinary collaboration among Digital Earth users as well as the creation of user-configured virtual research facilities/test-beds.
- Respond to Innovation: To integrate new scientific and technological derived paradigms in operational infrastructures in response to the latest Digital Earth requirements.
- Promote Virtualisation: To stimulate, educate and support the creation of virtual Digital Earth research communities.

http://www.genesi-dec.eu/project/

Comment:
The project is using two platforms: a development platform for definition, implementation and testing, and an operational platform for setup, deployment and operation.
Semantics is used for metadata, data, services annotation, and for search operation.
A security interoperability framework has been designed to support federation and interoperability between stakeholder security implementations.

6.19 GeoMind

Description:
GeoMind: Geophysical Multilingual Internet-Driven Information Center

An Internet-driven multilingual information system, integrating geophysical data coming from national data holdings, has been designed, developed and set up, developed also metadata and data representation standards, thus supporting further integration and dissemination of European or global geoinformation holdings. The system is addressing to local and central authorities, prospecting companies, research and education institutions and citizens. It enables the project partners to adapt to the pressures for continuous innovation and technological challenge, due to enlargement and completion of European market and intensification of competition in the field of services based on geological, geophysical and environmental data.

http://www.geomind.eu/portal/home.jsf

Comment:
Two main components are of interest: the portal and the multi-lingual metadata catalogue with its search engine used in an international context.

6.20 GeoMol

Description:
To meet the EU's ambitious targets for carbon emission reduction, renewable energy production must strongly be upgraded and made more efficient and capable for grid energy storage. Alpine Foreland Basins feature a unique geological inventory which can contribute substantially to tackle these challenges. Deep 'Molasse' basins at the fringe of the Alpine mountain range offer an abundant geothermal potential and storage capacity for weather-dependent wind and solar energy, as well as space for underground storage of gas or CO2. Exploiting these natural subsurface resources (geo-potentials) will strongly compete with existing oil and gas claims and groundwater issues. Thus, the efficient use and sustainable management of geo-potentials requires a holistic and transnational approach. In order to serve transnational decision-making, GeoMol will provide consistent 3-dimensional subsurface information based on coherent evaluation methods and commonly developed criteria and guidelines.

http://www.alpine-space.eu/projects/projects/detail/GeoMol/show/

Comment:
The project is starting on October 2012, then no results are yet available but it should be analysed next year during the architecture specification phase, especially for the management of 3D models.

6.21 Geo-Seas

Description:
Geo-Seas is implementing an e-infrastructure of 26 marine geological and geophysical data centres, located in 17 European maritime countries. Users are enabled to identify, locate and access pan-European, harmonised and federated marine geological and geophysical datasets and derived data products held by the data centres through a single common data portal.
Common data standards and exchange formats are agreed and implemented across the data centres. In addition to the data delivery activities, the project has also developed a number of viewing services for seismic and lithological log data.
A 3D viewing software tool has been developed on the basis of NASA World Wind for viewing bathymetric DTMs.

http://www.geo-seas.eu/

Comment:
The project enables the connection of several data centres and services providers based on the use of standard specifications. Access is provided by the project portal.
It has developed a metadata catalogue (metadata format was updated for geophysical data), common vocabularies (adaptation of SeaDataNet components)
Several tools have been developed for viewing high resolution seismic reflection data and analysis; viewing seabed maps and geological logs, and a 3D Viewer.

6.22 GEOWOW

Description:
GEOWOW (GEOSS Interoperability for Weather Ocean and Water) will improve the functionality and performance of the GEOSS Common Infrastructure (GCI), the project will address the following objectives:

- Further consolidating global data discovery and enabling global access to, and use of, Earth Observation data and resources (computing, data handling tools, models…) through the GCI.
- Developing tools and protocols promoting the implementation of the GEOSS Data Sharing Principles, and the re-use and dissemination of Earth Observation data, whilst addressing identified concerns expressed by data producers.
- Developing the operational capabilities of the GCI through applications in three areas:
    - Weather, with a focus on unified access to EO and forecasting systems for hazard and extreme meteorological events.
    - Water, with a focus on hydrological applications and run-off process using in situ and satellite data.
    - Ecosystem, with a focus on the implementation of GOOS by engineering and testing access to Ocean data via the GCI.

- Enhancing multidisciplinary interoperability.
- Analysing the benefits of GEOSS for Europe using models linking economy, environment, and society.

http://www.geowow.eu/index.html

Comments:
The GEOWOW project will take into account the results of EuroGEOSS and GENESI-DEC projects.
A work package is in charge of the architectural design, and of the development of its components, upgrading the current CGI components to the new standards and technical recommendations emerged from OGC, GEOSS Pilots, INSPIRE, GENESI-DR project. The work package will also integrate these components (discovery, access, and use) into the GEO portal.
Another objective is for conceptualizing and putting forward the multidisciplinary interoperability attributes by examining and suggesting the bonds between the thematic areas, the existing technologies and standards and the system architecture.

6.23 GS Soil

Description:
The GS Soil project aimed to establish a network to improve the access to spatial soil data in Europe. The project pursued the following objectives:

- to support of the development of a European spatial data infrastructure for soil data by improving the accessibility of digital soil data.
- to develop methods to produce interoperable spatial soil data, and the analyse of requirements to harmonise soil information.
- to develop of "harmonised" metadata and spatial data sets.
- to establish and operate a network of services provided on basis of a well operating project data infrastructure.

http://gssoil-portal.eu/ingrid-portal/portal/cms-info/cms-info-1.psml

Comment:
The architecture is very similar to the OneGeology-Europe architecture with a portal, a metadata catalogue with possible cascading on existing catalogues (metadata are based on INSPIRE requirements), a multilingual thesaurus specific for soil, and web services for viewing and downloading.
A security framework to secure access of WMS/WFS (access policies or click-through licences) has been designed.

6.24 Helix Nebula

Description:
This high-level European event organised by the Helix Nebula consortium is part of an overall cloud computing initiative which aims to build a cloud computing platform offering a unique resource to governments, businesses and citizens. Participants will have the unique opportunity to interact with cloud computing experts, SMEs, representatives of the European Commission and leading research centres. Each participant will actively contribute to shaping the overall cloud strategy and provide feedback and suggestions on the results achieved during the proof of concept stage.
The partnership brings together leading IT providers and three of Europe's leading research centres, CERN, EMBL and ESA in order to provide computing capacity and services that can elastically meet big science's growing demand for computing power and cross-domain community building.

http://helix-nebula.eu/

Comment:

ESA and CERN have defined a European Industrial Strategy for a Scientific Cloud Computing Infrastructure to be implemented by 2020.
The first goal of this strategy is to establish a cloud computing infrastructure for the European research area serving as a platform for innovation and evolution of the overall infrastructure.
Technical, legal, and procedural issues will be addressed by the Helix Nebula project, and this will be of interest for EGDI as the possible use of cloud computing is in the scope of EGDI design architecture task.

6.25 iCordi

Description:
iCordi (International collaboration on research data infrastructure) focuses on coordinating a series of cross-infrastructure experiments on global interoperability with a selected group of projects and communities. Each prototype addresses a specific community driven use case identifying best-of-breed solutions and the remaining challenges.

Earth Sciences:
iCORDI has particularly strong links into large, international Earth science collaborations, from both its core partners and through the wider EUDAT network. Earth and ocean sciences are well represented through CNR and Athena. The European Plate Observing System (EPOS) and the European Network for Earth Systems Modelling (ENES) are EUDAT

core communities, with EMSO, IAGOS-ERI and ICOS already connected as associate communities.
ICORDI has gathered a portfolio of five proposed prototype activities, covering seismological data processing, marine data interoperability, hydro-meteorological simulation infrastructure and international standards activities.
The effort will be focused on two different dimensions: vertically, demonstrating a federated access to distributed datasets within the seismology domain capitalizing EUDAT results; horizontally, debating on technology solutions to implement a cross-domain, robust and reliable infrastructure.

http://www.icordi.eu/

Comment:
Within the prototype activities, one is about infrastructure and international standards. For the common metadata, and common infrastructure iCORDI is connected with COOPEUS initiative and EUDAT to agree on data and metadata standards. There is also a connection between EPOS and EarthCube to better understand physical processes by integrating heterogeneous data.
The interest for EGDI is to understand how these connections operate to define the best relationship between EGDI and these research infrastructures.

6.26 InGeoCloudS - INspired GEOdata CLOUD Services

Description:
The INspired GEOdata CLOUD Services (InGeoCLOUDS) project aims at demonstrating the feasibility of employing a cloud-based infrastructure coupled with the necessary services to provide seamless access to geospatial public sector information, especially targeting the geological, geophysical and other geoscientific information. This kind of information possesses interesting characteristics like the size of the available data, the existing metadata descriptions (mostly according to the European Directive INSPIRE) and the current availability of related services that can be moved to the cloud.

http://www.ingeoclouds.eu/

Comment:
The interest for EGDI are the guidelines InGeoClouds will provide to move services and data to the cloud. The project will define requirements of the underlying infrastructure, and for the services.

6.27 LESSLOSS

Description:
LESSLOSS project addresses research issues on seismic engineering, earthquake risk and impact assessment, landslides monitoring, mapping and management strategies, improved disaster preparedness and mitigation of geotechnical hazards, development of advanced methods for risk assessment, methods of appraising environmental quality and relevant pre-normative research

http://www.lessloss.org/main/index.php

Comment:
This project has not developed an infrastructure but some tools (based on GIS methodologies, for landslide monitoring for example) that might be taken into account to be connected to the EGDI.

6.28 NERA – Earthquake risk assessment and mitigation

Description:
Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation.
NERA (2010-2014) is an EC infrastructure project that integrates key research infrastructures in Europe for monitoring earthquakes and assessing their hazard and risk.
NERA integrates and facilitates the use of these infrastructures and access to data for research, provides services and access to earthquake data and parameters, and hazard and risk products and tools.
NERA activities are coordinated with other relevant EC-projects and European initiatives, and contribute to the ESFRI EPOS infrastructure and the OECD GEM (Global Earthquake Model) program.

Portal and services:
The primary goals are to improve discoverability, access, and usability of the data and products from the various NERA activities by advancing the Earthquake Data Portal and supporting web services to provide a complete framework for the access and distribution of NERA project data sets and results.

http://www.nera-eu.org/ (NERIES, previous project. http://www.neries-eu.org/)

Comment:
To improve discoverability, access, and usability of the data and products, the portal and supporting web services provide a complete framework for the access and distribution of data sets and results.
The web data services layer is extended and standardized within a common services architecture. The common services architecture will enable the development of higher-order client applications that access the data provider services directly.
WMS and WFS are based on Geoserver and should be visible in the OneGeology portal.

6.29 Ocean Data Interoperability Platform (ODIP)

Description:
The objectives of ODIP are:

- to establish an EU/USA/Australia/IOC-IODE co-ordination platform to facilitate the development interoperability between these regional ocean and marine data management infrastructures
- to demonstrate this coordination through the development of several joint EU-USA-Australia prototypes that would ensure persistent availability and effective sharing of data across scientific domains, organisations and national boundaries.

Not yet a web site available, but a Cordis link:
http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=13380280

Comment:
The project is just starting (October 2012), then no results are yet available, but should be analysed next year for the architecture specification phase.

6.30 OneGeology-Europe

Description:
OneGeology-Europe aims to create dynamic digital geological map data for Europe. It will make a significant contribution to the progress of INSPIRE - i.e. develop systems and protocols to better enable the discovery, viewing, downloading and sharing of core European spatial geological data.
OneGeology-Europe addresses licensing and multi-lingual aspects of sharing geological knowledge and demonstrates best practice examples of the delivery and application of geological spatial data in the public and private sectors.
20 European nations were represented, so the geographical coverage is quite important.

http://www.onegeology-europe.org

Comment:
OneGeology-Europe has implemented all technical chapters of INSPIRE: metadata, data, services, data sharing and monitoring. As not all INSPIRE rules were completed at the end of the project (October 2010) the implementation is not 100% INSPIRE compliant but very close to what is/will be required.
From an infrastructure point of view, OneGeology-Europe is a good example of a starting point for EGDI.

6.31 OpenQuake / GEM (Global Earthquake model)

Description:

GEM is a global collaborative effort with the aim to provide organisations and people with tools and resources for transparent assessment of earthquake risk anywhere in the world. By pooling data, knowledge and people, GEM acts as an international forum for collaboration and exchange, and leverages the knowledge of leading experts for the benefit of society
OpenQuake is a dynamic platform to serve as a true hub for earthquake risk assessment in all areas of the world, used from global to local levels. Users support each other in analysis and planning and hereby increasingly contribute to making society more resilient to earthquakes.
The main functions are:

- Calculate:

The OpenQuake platform integrates open-source applications with homogenized data and models, allowing users to model seismic hazard and risk transparently and according to the latest science. From ground motion fields to hazard spectra to maps of estimated human loss and mean economic loss; users will be able to produce a great variety of custom outputs by combining own data and (local) knowledge with GEM products.

- Explore:

The platform leverages upon open-source geospatial technologies [Geonode, QuantumGIS, etc.] to allow users to work in an intuitive GIS-environment. Users will explore earthquake hazard and risk by interacting with dynamic maps, indicators and graphs, develop their own maps, but also capture and integrate new data. To help users understand risk better and to facilitate risk management, decision support tools are also part of the platform.

- Share:

Sharing of data and risk information, best practice and approaches is key to assessing risk better. The platform is to serve as a clearinghouse for all those critical in- and outputs. It will link users from around the globe so they can work together to assess risk.

The OpenQuake Engine:
The OpenQuake Engine is an open-source application for performing seismic hazard and risk analysis. The Engine is built on an assorted stack of open-source software components, including PostgreSQL, PostGIS, Celery, RabbitMQ, Redis, Numpy, and Scipy. The

core seismic hazard calculation performed by the OpenQuake Engine uses the OpenSHA Java library, which is licensed under Apache v2.0.

http://beta.globalquakemodel.org/

Comment:
The platform deals with global homogenized datasets used in GEM models on seismic hazard, exposure, …Tools are developed for users to interact with the data.

6.32 PanGeo

Description:
PanGeo provides information about the stability of the ground on which we live, work and play. Ground instability, or 'geohazards', affect us all. They can be dangerous and costly, yet information on these phenomena can be difficult if not impossible to obtain. The PanGeo service provides entirely free access to geohazard information for many of the largest cities in Europe
The PanGeo service is aimed at local authorities, civil protection agencies, geological surveys, the public, insurers and businesses providing environmental and land reporting services. The service and products available are entirely free to use, even for commercial gain, providing appropriate credits are given (see licence terms).
PanGeo products are made from integrations of:

- Satellite measurements of ground and building movement
- Geological information already held by National Geological Surveys
- The landuse and landcover information contained within the EC's Urban Atlas
- Urban Atlas

Products are provided in real-time from the servers of the 27 EU National Geological Surveys, via the PanGeo Portal.

http://www.pangeoproject.eu/

Comment:
Infrastructure is based on OneGeology-Europe. The portal will deliver a Ground Stability Layer for each of the 52 towns for which PanGeo information is available. For each town a geohazard summary (a document explaining the geological interpretation) is available.
The ground stability polygons are compliant with INSPIRE data specifications for Natural Risk Zones.

6.33 Planet Data

Description:
PlanetData aims to establish a sustainable European community of researchers that supports organizations in exposing their data in new and useful ways. The ability to effectively and efficiently make sense out of the enormous amounts of data continuously published online, including data streams, (micro)blog posts, digital archives, eScience resources, public sector data sets, and the Linked Open Data Cloud, is a crucial ingredient for Europe's transition to a knowledge society. It allows businesses, governments, communities and individuals to take decisions in an informed manner, ensuring competitive advantages, and general welfare.
The PlanetData project is based upon three objectives that together create a durable community made up of academic and industrial partners. This community is supported in conducting research in the large-scale data management area through the provision of data sets and access to tailored data management technology. This community also benefits from a comprehensive program of training, dissemination, standardization, and networking activities, intended to strengthen existing collaborations and establish new ones, to educate organizations in key questions related to open data exposure, and to transfer research results towards industry. PlanetData's objectives are:

- Research: To bring together approaches to large-scale data management from different disciplines in order to create holistic solutions to the challenges faced when dealing with planetary-scale data.
- Data provisioning and management: To provide software to support large-scale data provisioning, made available via the PlanetData Lab, supporting relational, graph, and stream processing, for researchers to test and validate their techniques. To create definitive vocabularies for the description of data sets and their context. To build a catalogue of data sets in vertical domains chosen for their high adoption potential and data management needs. To publicize guidelines and best practices for provisioning, such that available data sets can be more readily consumed by end-users and efficiently assembled into innovative products and services.
- Impact: To provide a medium through which the research results and empirical findings of the PlanetData network can be used to improve the education level related to large-scale data management in both academia and industry; to bring together researchers from disparate disciplines in order to form an integrated community that can support organizations in publishing their data in a way that is purposeful, thus addressing key challenges of large-scale data management; to encourage (industrial) uptake through standardization, and strategic dissemination and networking events.

http://planet-data.eu

Comment:
The project is focusing on large-scale data, but deals with many components or topics of interest for EGDI infrastructure, for example the need for non-technical issues as guidelines, training, networking, dissemination, … to go with the technical components.

6.34 ProMine

Description:
ProMine: Nano-particle products from new mineral resources in Europe
The non-energy extractive industry is a significant contributor to the economy of the EU providing metalliferous and non-metalliferous mineral resources to the society as well as direct and indirect employment. The philosophy behind ProMine is to stimulate the extractive industry to deliver new products to manufacturing industry. ProMine will kick-start a process of renewal whose momentum will carry over into the coming decennia.

Main objectives of the project:
- To develop the first pan-European GIS-based database containing the known and predicted metalliferous and non-metalliferous resources, which together define the strategic reserves (including secondary resources) of the EU.
- To calculate the volumes of potentially strategic metals (e.g. cobalt, niobium, vanadium, antimony, platinum group elements and REE) and minerals that are currently not extracted in Europe.
- To develop five new, high value, mineral-based (nano) products.
- To enlarge the number of profitable potential targets in Europe.
- To establish a new, cross-platform information group between the European Technology Platform on Sustainable Mineral Resources (ETP-SMR) and other platforms

http://promine.gtk.fi/

Comment:
ProMine delivers two central databases (mineral deposits, and anthropic concentrations) populated by data collected from several European countries. On top of these databases are setup WMS services and WFS services (both INSPIRE compliant) to deliver data according to the INSPIRE Data Specifications for Mineral Resources.

6.35 SAFER – Seismic early warning for Europe

Description:
The SAFER project aims to fully exploit the possibilities offered by the real-time analysis of the signals coming from seismic networks for a wide range of actions, performed in a time interval of a few seconds to some tens of minutes. These actions range from the shutting

down of critical systems of lifelines, industries, highways, railways, etc., the activation of control systems for the protection of crucial structures, the supplying of information to support decision making for the rapid response of emergency management services (e.g. ground shaking maps, continuously updated damage scenario estimates, aftershocks hazard etc.).

SAFER is not only focused on research but also aims to disseminate the results, by providing Civil Protections Authorities with tools for crisis management. The project develops improved algorithms for the rapid determination of earthquake source parameters, elaborating new concepts for providing real-time alert maps and predicted shake maps. Development of fast algorithms for damage scenario simulations and the deployment of real time structural control mechanism for the immediate protection of threatened infrastructure are also considered within the project. The tools will be tested and applied to several test cities: Istanbul, Bucharest, Athens, Napoli and Cairo

http://www.saferproject.net/index.htm

Comment:
The project provides tools to process real time data for crisis management. There is no specific infrastructure designed, except to integrate seismic sensors, to monitor devices with real time reporting and early warning capabilities.

6.36 SafeLand

Description:
SafeLand has the objectives to (1) provide policy-makers, public administrators, researchers, scientists, educators and other stakeholders with an improved harmonised framework and methodology for the assessment and quantification of landslide risk in Europe's regions; (2) evaluate the changes in risk pattern caused by climate change, human activity and policy changes; and (3) provide guidelines for choosing the most appropriate risk management strategies, including risk mitigation and prevention measures.

http://www.safeland-fp7.eu/

Comment:
The project provides guidance for interoperability and harmonisation of landslide databases in agreement with the EU Soil Thematic Strategy and its associated Proposal for a Soil Framework Directive, and for achieving interoperability and harmonization in agreement with INSPIRE Directive.

A software for early-warning based on real-time data was developed to support technical staff in data analysis and the decisional process (real-time analysis and evaluation of geo-scientific monitoring data, including threshold evaluation). The prototype of a web-based "toolbox" of innovative and technically appropriate prevention and mitigation measures does a preliminary assessment and ranking of up to 60 structural and non-structural landslide risk mitigation options.

6.37 SciDIP-ES

Description:
SciDIP-ES: (Science Data Infrastructure for Preservation – with focus on Earth Science)
The objective of SciDIP-ES is to develop a European framework for the long term preservation of Earth Science data through the definition of common preservation policies, the harmonisation of metadata and semantics, and the deployment of generic data preservation services

http://www.scidip-es.eu/scidip-es/

Comment:
This project addresses a topic, the data preservation, that should be taken into account for EGDI.

6.38 SHARE – Seismic Hazard Harmonization in Europe

Description:

SHARE's main objective is to provide a community-based seismic hazard model for the Euro-Mediterranean region with update mechanisms. The project aims to establish new standards in Probabilistic Seismic Hazard Assessment (PSHA) practice by a close cooperation of leading European geologists, seismologists and engineers.

http://www.share-eu.org/

Comment:
The project provides a computational infrastructure for seismic hazard assessment and produce a probabilistic seismic hazard assessment model in close cooperation with GEM (Global Earthquake Model).
4 types of services have been setup: view service, data access service, hazard access service, and computation access service.

6.39 SUBCOAST

Description:
SubCoast: Assessing and monitoring subsidence hazards in coastal lodlanf around Europe.

SubCoast aims at developing a GMES-downstream service for assessing and monitoring subsidence hazards in coastal lowland areas around Europe. SubCoast develops GMES-downstream services based on satellite data, in-situ measurements and geoscientific models. SubCoast builds upon ESA's GMES Service Element Terrafirma which provides a ground motion hazard information service.

http://www.subcoast.eu/

Comment:
The project has developed a "Data Provision Service" built incorporating links to the existing GMES Core Services (Geoland2, MyOcena), GMES Service Element Terrafirma and auxiliary data-streams consisting of terrestrial data.

6.40 Terrafirma

Description:
Terrafirma, is an ESA GMES project aimed at providing civil protection agencies, disaster management organisms, and coastal, rail and motorway authorities with support in the process of risk assessment and mitigation by using the latest technology to measure terrain motion from satellite radar data. Using these data acquired 800km above the Earth's surface, and in conjunction with expert interpretation by national geoscience organisations, the project provides a pan-European ground motion hazard information service in each of the 27 member states of the EU
The third phase of Terrafirma focuses on several thematic lines for terrain motion analysis: tectonics, coastal lowland subsidence & flood defence, hydrogeology (ground water issues, landslides, inactive mines), Wide Area Mapping service.

http://www.terrafirma.eu.com/

Comment:
Terrafirma provides two types of product: Advanced Terrain Motion Mapping and Advanced Terrain Motion Modelling, available for a number of application themes.
The products are specified (databases, layers, reports, images, …)

6.41 ThermoMap

Description:
The ThermoMap project (Area mapping of superficial geothermic resources by soil and groundwater data) focuses on the mapping of very shallow geothermal energy potentials in Europe.Geothermal energy.
The ThermoMap project will harmonise and analyse already existing data collections (geological, hydrogeological, soil, climate and relief geodata) with standardised methods to calculate a value for the geothermal potential on three different low depth levels in order to

help finding favorable areas for superficial geothermal exploitation in a very short time and without high costs.

The resulting geothermal potential values will be integrated in an Open Source WebGIS as well as all necessary geodata. The 12 participating project partners of 9 EU member states defined one or two test sites in each country. For these 14 Test Areas the geodata and calculated geothermal potential values will be shown on cadastral parcel level, while for the entire EU area there will be created a shallow geothermal potential Outline Map in scale 1 : 250,000.

http://www.thermomap-project.eu/

Comment:
The project does not design an infrastructure. The data will be processed by a GIS tool and the results will be delivered through a Web GIS system. The necessary map layers are provided by WMS services..

6.42 VERCE

Description:
VERCE: Virtual Earthquake and seismology Research Community in Europe e-science environment.
The objectives of the VERCE project are to:

- Provide to the Virtual Earthquake and seismology Research Community in Europe, a data-intensive e-Science environment - based on a service-oriented architecture - integrating a number of specialized services, tools, data-flow and work-flow engines, to support the data-intensive applications of this community and beyond to the EPOS community.
- Provide a framework wrapping the seismology data-infrastructure resources and services with a set of distributed data-aware Grid, Cloud and HPC resources provided by the European e-infrastructures and the community.
- Produce a core of pilot data-intensive applications and use cases of the Virtual Earthquake and seismology Community of research in Europe that exemplify the power of the platform architecture and its capabilities.
- Deliver a scientific gateway providing a unified access, management and monitoring of the platform services and tools, domain specific interfaces supporting the coevolution of research practices and their supporting software.
- Deliver an 'intellectual ramp' providing safe and supported means for researchers and users of the community at large to advance to more sophisticated data use through tailored interfaces and facilitators integrated within the scientific gateway.
- Deliver a 'research-methods ramp' through a toolkit of training programs for data intensive research - composed as a set of training session material, demonstrators, and best practice guides - based on tailored use-case scenarios and produce data-intensive applications of the community.

- Provide a collaborative environment between the earthquake and seismology research community and the computer scientists, system architects and data-aware engineers, fostering the emergence of 'research technologists' with sustained mastery for data-handling methods and a thorough understanding of the research goals.

Comment:
The interest or this Research Infrastructure for EGDI is the service-oriented architecture with an efficient communication layer between the Data and the Grid infrastructures, and HPC. It will address also workflow and data sharing mechanisms for coupling HTC data analysis and HPC data modelling applications.

6.43 WMO Information System

Description:
The World Meteorological Organisation Information System (WIS) is the pillar of the WMO strategy for managing and moving weather, water and climate information in the 21st century. WIS provides an integrated approach suitable for all WMO Programmes to meet the requirements for routine collection and automated dissemination of observed data and products, as well as data discovery, access and retrieval services for all weather, climate, water and related data produced by centres and Member countries in the framework of any WMO Programme.
WIS is being designed to dramatically extend WMO Members' ability to collect and disseminate data and products. It will be the core information system utilized by WMO Members, providing linkages for all WMO and supported programmes associated with weather, climate, water, and related natural disasters.

http://www.wmo.int/pages/themes/wis/index_en.html

Comment:
The system is built on existing information systems with an extension of the information services through flexible data discovery, access and retrieval services to authorized users, as well as flexible timely delivery services. It will be implemented essentially through the Internet.
Three core elements are integrated: (1) a global observation system providing observational data for use in operational and research work, (2) Integrated networks of telecommunication facilities for collection and distribution of observational data and processed information, (3) Centres (at various geographical levels) to provide processed data, analysis and forecast products.

7 Initiatives

7.1 EarthCube

Description:

NSF (National Science Foundation) seeks transformative concepts and approaches to create integrated data management infrastructures across the Geosciences. In a new partnership, the Geosciences Directorate (GEO) and the Office of Cyberinfrastructure (OCI) recognize the multifaceted challenges of modern, data-intensive science and education and envision an environment where low adoption thresholds and new capabilities act together to greatly increase the productivity and capability of researchers and educators working at the frontiers of Earth system science.

GEO continues to make substantial investments in collecting data through NSF-supported research facilities and projects, and in helping the geosciences community utilize data collected by other entities around the world. Similarly, OCI makes substantial investments in advanced high-performance computing, data infrastructure, software development, virtual organizations and networking. It is time to integrate these data and technologies in an open, adaptable and sustainable framework (an "Earth-Cube") to enable transformative research and education in Earth System Science; foster common data models and data-focused methodologies; develop next generation search and data tools; and advance application software to integrate data from various sources and advance knowledge.

http://www.nsf.gov/geo/earthcube/

Comment:

The interest for EGDI is to understand how EarthCube will integrate many available data and technologies in a network, as EGDI should be a part of this network.

Several white papers are under developments, and should be analysed when available. The description (from the EarthCube web site) is below:

> **Governance:**
> Tools, processes & mentoring to enable & manage cross-discipline, cross-sector, cross-jurisdiction, long-term, adaptive collaboration. Processes include patterns for describing complex systems, forming and managing working groups, developing compliance guidelines, and advancing the design lifecycle of relevant standards and best practices. Must make effective consideration of social, legal, and other institutional issues, including trust, uncertainty, authentication/authorization for discovery and access. Consideration of GEOSS approaches.
>
> **Science scenarios**
> Motivational vignettes describing data-intensive, cross-disciplinary research that illustrates the need for an open, adaptable and integrative cyberinfrastructure. These sce-

narios will cover a range of topics in the Geosciences and target current and anticipated future problems faced by researchers, and their support systems (research centers, libraries, data centers and archives and commercial service providers) working to address these challenges.

Cyber-architecture for Science
Achieving the challenges in geoscience research requires innovation and paradigm shifts in cyberinfrastructure. Information technology must advance to meet the emerging approaches to science, e.g., Four Paradigms of science epistemology; Data-intensive systems; Multi-disciplinary collaborations to address complexity. A cyber-architecture identifies the repeatable patterns, reusable components, and open standards that provide starting point for innovative developments.

Data interactive publications :
The overall objective is simple: make it possible for authors to create online publications that enable readers to access, analyze, and display the data discussed in the publication.

Sensor Webs
Proposed solution to advance existing NSF funded capabilities to build an Integrated Sensor Web System, supporting researches familiar with the sensor data and allowing others to discover and use sensor data beyond their original research intent.

7.2 EPOS – European Plate Observing System

Description:
The European Plate Observing System (EPOS) is the integrated solid Earth Sciences research infrastructure approved by the European Strategy Forum on Research Infrastructures (ESFRI) and included in the ESFRI Roadmap. EPOS is a long-term integration plan of national existing RIs.
The goal of EPOS is to promote and make possible innovative approaches for a better understanding of the physical processes controlling earthquakes, volcanic eruptions, unrest episodes and tsunamis as well as those driving tectonics and Earth surface dynamics. Integration of the existing national and trans-national RIs will increase access and use of the multidisciplinary data recorded by the solid Earth monitoring networks, acquired in laboratory experiments and/or produced by computational simulations. Establishment of EPOS will foster worldwide interoperability in Earth Sciences and provide services to a broad community of users
EPOS will ensure secure storage of geophysical and geological data providing the continued commitment needed for long-term observation of the Earth.

The e-infrastructure
The e-infrastructure requires development of new data standards, and integration of distributed archives, computing facilities and experimental facilities exploiting GRID/HPC

technology and advanced portal technologies. An efficient and pragmatic approach must be developed in the preparatory phase of EPOS through efficient interaction between geoscientists, the user community and skilled IT specialists. The range of preferably Web-based tools and services to be developed, includes:

- Web-based portal solutions for data distribution, data mining and data archiving
- Grid-based structures to enhance computing power
- Standardized protocols or protocol services for data exchange and data availability
- Web based services for the integration of new sensors, instruments, observatories or labs into the infrastructure
- Repositories for data assimilation tools, modelling tools and data visualization tools
- Repositories for observational and experimental raw data, pre-processed data products and modelling/simulation data including the respective metadata.

The EPOS infrastructure is based on existing, discipline-oriented, national data centres and data providers managed by national communities for data archiving and mining. EPOS will work to harmonize these national data centres, promoting data authentication and standardization as well as shared national implementation plans. These national research infrastructures will be integrated during the preparatory phase to form EPOS Data Centres, visible as discipline service providers (seismological data, GPS geodetic data, geological repositories, volcano observatory data, etc...).

Each EPOS Data Centre will have its own computational facilities. This will allow the establishment of effective distributed data storage and mining as well as providing essential computational resources that will contribute to high-performance computing facilities and grid initiatives at European level. This will facilitate the subsequent step in the e-infrastructure integration plan, which consists of the establishment of the EPOS Core Services representing the EPOS Data services serving the multiple communities integrated by EPOS.

The EPOS Core Services will provide the top-level service to users including access to multidisciplinary data and metadata, virtual data from modelling and solid Earth simulations, data processing and visualization tools as well as access to high-performance computational facilities. EPOS will enhance data processing and modelling capacity and capability as well as develop new theoretical and numerical tools to harness computational power in a distributed European architecture.

http://www.epos-eu.org/

Comment:
EPOS Core services with two layers built on top of the “national layer” (national research infrastructures and data centres):

- Integration layer (data oriented) with:
    - Discovery functions
    - Portal functions
    - Access to high performance computing (HPC)
    - Access to data products

    - Data mining
    - Visualisation tools
    - Access to modelling tools
    - Processing tools
    - Training tools and tutorials
- Community layer (thematic services) with:
    - Hazard maps
    - Natural laboratories
    - Supersites
    - Experimental and analogue laboratories
    - Seismological thematic centres
    - Geomagnetic thematic centres
    - Geodesy
    - Satellite data and information
    - Volcano observatories
    - Geological repositories

The relationship between EGDI and EPOS have to be well specified, as EPOS represents the "world" of Research Infrastructures that EGDI would like to communicate with. The thematic domain (geosciences) is shared by the two projects, and many issues to address are the same. Therefore a strong link with EPOS must operate during the next phase of WP4 (the technical design).

7.3 EUDAT – European Data Infrastructure

Description:
The EUDAT vision is to support a Collaborative Data Infrastructure which will allow researchers to share data within and between communities and enable them to carry out their research effectively. EUDAT aims to provide a solution that will be affordable, trustworthy, robust, persistent and easy to use.
Mission:

- Help fulfill the vision of a European Data e-infrastructure by providing a sustainable platform of technologies, tools and services driven by user needs.
- Engage users (including individual researchers along with representatives from universities, research labs, and libraries) in defining and shaping a platform for shared services that makes it possible for data-intensive research to span all the scientific disciplines.
- Produce the common low-level services that are required to provide the level of interoperation and trust of data that is necessary to support both widespread access to data, and the long-term preservation of data for use and re-use.
- Ensure that the data infrastructure is sufficiently robust to keep pace with the expected acceleration of the scale and complexity of scientific data being generated within the ERA and beyond.

http://www.eudat.eu/

Comment:
EUDAT has defined a collaborative data infrastructure with two layers:

- Common data services : persistent storage, identification, authenticity, workflow execution, …
- Common support services : data discovery & navigation, workflow generation, annotation, interoperability

The users and data providers use these two groups of services. Trust and data curation have to be taken into account across all services.

7.4 Eye on Earth

Description:

Eye on Earth is a global public information network, launched in May 2008 in the form of a Public Private Partnership by the European Environment Agency (EEA), designed to share environmentally relevant data and information through interactive map-based visualisations and give access to data based on GIS.
Eye on Earth also contributes to increasing awareness about environmental issues. Based on a model of crowdsourcing, this virtual cloud allows users to visualise, create, share, collect and interact with maps, bringing together expertise from public institutions, industry, international organisations, NGOs and private users.

Eye on Earth is shaping a new profile of citizen in this domain. Citizens, mainly scientists, contribute by providing their own ratings of air and water quality, creating new custom maps and sharing them with other users. Contributions can be uploaded both through web and mobile applications.

One specific feature of Eye on Earth is the "Watches", which are key components that aim to collect and compile environmental data from diverse sources and transform them into relevant information.

http://www.eyeonearth.org/en-us/Pages/Home.aspx

Comment:
Eye on Earth is a good example of crowdsourcing, to be taken into account if there is such a requirement for EGDI.

7.5 GEOSS

Description:

GEOSS: Global earth Observation System of Systems

The development of GEOSS has multiple strands as identified in its 10-year implementation plan: developments in the societal benefits areas (SBAs), harmonise and sustain earth observation and the use of modelling, data sharing, architecture and data, capacity building and outreach. A key component of GEOSS is the GEOSS Common Infrastructure (GCI) which (should) optimally allow users to discover, access and use the data and services available through GEOSS. Status of the GEOSS Common Infrastructure (GCI):
At present, the GCI includes four main components:
1. The GEO Portal provides the direct web interface through which the user accesses GEOSS and searches for information and services.
2. The GEOSS Clearinghouse connects to the various GEOSS components and services, collects and searches their information and distributes data and services via the Portal to the user.
3. The GEOSS Components and Services Registry is where the governments and organisations, that contribute components and services to GEOSS, provide essential details about the name, contents, and management of their contribution. This assists the Clearinghouse, and ultimately the user, to identify the GEOSS resources that may be of interest.
4. The GEOSS Standards and Interoperability Registry that enables contributors to GEOSS to configure their systems so that they can share information with other systems.

GEOSS Architecture
The GEOSS Architecture describes how components fit together to produce an overall system of systems capable of providing data and information that will better satisfy requirements than individual components or systems of which it is composed.

"The success of GEOSS will depend on data and information providers accepting and implementing a set of interoperability arrangements, including technical specifications for collecting, processing, storing, and disseminating shared data, metadata, and products." (from GEOSS Implementation plan.

http://www.earthobservations.org/index.shtml

GEOSS Architecture Implementation Pilot:
The **GEOSS Architecture Implementation Pilot (AIP)** develops and deploys new process and infrastructure components for the GEOSS Common Infrastructure (GCI) and the broader GEOSS architecture.

http://www.ogcnetwork.net/AIP

7.6 Google

Google is registered in this analysis not as a model for EGDI Architecture but more than an infrastructure EGDI should collaborate with to provide data, layers to the map viewer.

Another possible use is the cloud platform described below:
Google Cloud Platform allows to build applications, host websites, store data and analyze information using the same scalable systems that power Google.

- App Engine to build scalable web applications.
- Compute Engine to solve large-scale processing and analytic problems in the cloud using Google's virtual computing infrastructure.
- Cloud Storage to store data in the cloud with direct access to Google's storage and networking infrastructure as well as powerful authentication and sharing mechanisms.
- BigQuery to analyze massive datasets interactively.
- Cloud SQL to move the data, applications and services in and out of the cloud effortlessly using a familiar MySQL database.

https://cloud.google.com/

Comment:
EGDI should consider to use the Google infrastructure as a cloud, to use the tools offered by Google for mapping - Google as a viewer for our data (2D or 2.5D viewer with Google Earth) - for computing statistics.
As for Research Infrastructures, EGDI will have to specify the connections with Google.

7.7 US- NGDC - National Geophysical Data Center

Description:
The Mission of NOAA's National Geophysical Data Center (NGDC) is to provide long-term scientific data stewardship for the Nation's geophysical data, ensuring quality, integrity, and accessibility.
NGDC provides stewardship, products, and services for geophysical data from our Sun to Earth and Earth's sea floor and solid earth environment, including Earth observations from space.

http://www.ngdc.noaa.gov/

Comment:
This web site gives access to a wide range of data related to Earth. It is possible to find data by using interactive map services, or by searching a catalogue of metadata. The

metadata provides the link for downloading data. The web site delivers also many reports and publications.

7.8 OneGeology

Description:
The mission of OneGeology is to make web-accessible the best available geological map data worldwide at a scale of about 1: 1 million.
The project's aims are to:

- create dynamic digital geological map data for the world.
- make existing geological map data accessible in whatever digital format is available in each country. The target scale is 1:1 million but the project will be pragmatic and accept a range of scales and the best available data.
- transfer know-how to those who need it, adopting an approach that recognises that different nations have differing abilities to participate.
- the initiative is truly multilateral and multinational and will be carried out under the umbrella of several global organisations

http://onegeology.org/

Comment:
OneGeology architecture is based on service oriented architecture (SOA) principles. The geological maps providers define metadata and setup a WMS, and when possible a WFS delivering information about lithology and age. The metadata are registered in a catalogue, accessed by the OneGeology portal to view the maps.
OneGeology is a good example of interoperability requesting a simple level of technical knowledge (define metadata and setup a WMS) to participate to the global geological map. OneGeology-Europe project used this architecture to improve it and made it more compliant with INSPIRE requirements.

7.9 ORFEUS - Observatories for European Seismology

Description:
ORFEUS (Observatories and Research Facilities for EUropean Seismology), founded in 1987, is the non-profit foundation that aims at co-ordinating and promoting digital, broadband (BB) seismology in the European-Mediterranean area.
Its activities are distributed between the ORFEUS Data Centre (ODC), gathering, archiving and providing waveform data, and four working groups, coordinating data availability and relevant developments
ORFEUS currently coordinates archiving of and access to earthquake waveform data from seismic stations in the European Mediterranean region through a European Integrated Waveform Data Archive in Europe. ORFEUS coordinated the EC Infrastructure (I3) project

NERIES (Network of Research Infrastructures for European Seismology) and is currently managing its successor, the EC Infrastructure project NERA.

http://www.orfeus-eu.org/

NERA: Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation
http://www.nera-eu.org/

Comment:
One task of ORFEUS (of the working group 1 - Observatory coordination of broadband data acquisition and data exchange) is to promote common standards and coordination among seismological networks. Another one is to promote the availability of seismological software. The web site gives access to earthquakes data, and make them viewable on Google maps.

7.10 US-GIN: Geoscience Information Network

Description:
USGIN is a conceptual framework for sharing geoscience data, a distributed data-sharing network, a collection of open-source applications, standards, procedures, and protocols for sharing geoscience data web-based, distributed, open-source, and interoperable.
USGIN is a distributed data-sharing network that uses open-source software and existing World Wide Web infrastructure and browsers. Anyone can access the data shared over USGIN, and any data that is shared according to USGIN specifications is automatically a part of the network.

http://usgin.org/

Comment:
One objective of USGIN is to give public access to interoperable digital geoscience data, meaning that data is web-accessible and discoverable. To satisfy this, USGIN provides data provides with specifications, best practices, workflows, and software applications.

The web site USGIN Lab provides technical information for the implementation
http://lab.usgin.org/about-labusginorg

7.11 USGS – SDI

Description:
To setup its SDI, USGS wrote a report describing the roadmap: “A Spatial Data Infrastructure Roadmap for the U.S. Geological Survey” National Academy of Sciences (2012).

Comment:
There is no established, validated process for developing an SDI. But 3 phases are suggested: preparation and planning; design, development and testing; rollout and refinement.

USGS suggests criteria for a successful implementation of an SDI that depends on an agency's roadmap and strategy, organizational leadership and culture, standardization, technical competence, funding and contracting, workforce competence, and cooperation and partnership.
And some more recommendations:

- To collaborate for developing an improving data standards (in accordance with international standards),
- To have a well-developed roadmap based on scientific and business practices,
- Technology and data standards are key to enabling interoperability of resources throughout a broad community,
- To invoke standardized quality requirements at the data sources,
- Importance of semantics and ontology in an SDI,
- Training is critical for introducing and maintaining an SDI,
- The SDI should serve as a powerful web platform for analysis and modeling: data from multiple sources to be integrated and linked to a unified model of a given system,
- To avoid moving massive amounts of data around, an SDI enables computations to be pushed as close to the data sources as possible,
- Documenting provenance is particularly important in the publication.

8 Directives, framework documents

8.1 GMES

Description:

GMES (Global Monitoring for Environment and Security) is a European system for monitoring the Earth.
GMES consists of a complex set of systems which collect data from multiple sources: earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors. It processes these data and provides users with reliable and up-to-date information through a set of services related to environmental and security issues...
The services address six thematic areas: land, marine, atmosphere, climate change, emergency management and security. They support a wide range of applications, including environment protection, management of urban areas, regional and local planning, agriculture, forestry, fisheries, health, transport, climate change, sustainable development, civil protection and tourism.

In the Strategic Implementation Plan for the Land Monitoring Core Services:

In order to mesh the individual components into a fully functioning service within a distributed architecture, two approaches, top-down and bottom-up, have to come together and a set of common rules agreed upon: which address the following issues:

- Standards & Specifications:

A service-oriented architecture approach based on interoperable service components needs to define component interfaces, capable of ensuring interoperability both, with other GMES services as well as internally between components and core and downstream elements. Internationally agreed standards (e.g. ISO, CEN) and open specifications (e.g. W3C, OGC) expressed in widely accepted languages (e.g. XML, UML) provide a sound framework for interface definitions.

- Interoperability:

Here at least the data harmonisation component will be addressed along with service interoperability in the strict sense at service level. Both aspects will take into account the ongoing work for the INSPIRE Implementing Rules (IR) under development. IRs are being drafted for data specifications and network services. Ortho-imagery and landcover/ landuse are part of the themes listed in the annexes of the INSPIRE Directive. High-level harmonisation aspects are available in a first draft, but we try to envisage LMCS as a unique opportunity to elaborate further and/or test the theme-specific part of data harmonisation. In parallel, the INSPIRE network service IRs focus on middleware service layers and their necessary interfacing. Here as well an LMCS testbed can emerge.

- Coordination and compliance checking:

In a distributed architecture which is expected to deliver operational services, an effective coordination mechanism is an absolute prerequisite for success. Given a service-oriented architecture, it will preferably be completed by a compliance checking and clearance activity.

http://www.gmes.info/pages-principales/overview/

Comment:

For EGDI the interest is to check how services, for other thematic domains than geology s.s., have been specified and connected to the distributed architecture. The three main issues addressed are of interest: the use of standards, the implementation of interoperability and harmonisation (with a strong link with INSPIRE requirements), and the coordination.

8.2 European Interoperability Framework (EIF)

Description:

The purpose of the European Interoperability Framework (EIF 2.0) from the "Interoperability Solutions for European Public Administrations" (ISA) program is:

- to promote and support the delivery of European public services by fostering cross-border and cross-sectoral interoperability;

- to guide public administrations in their work to provide European public services to businesses and citizens;
- to complement and tie together the various National Interoperability Frameworks
- (NIFs) at European level.

The EIF provides guidance to European public administrations as regards the definition, design and implementation of European public services. It introduces:

- 12 underlying principles summarising the expectations of public administrations, business and citizens regarding the delivery of public services;
- a conceptual model for public services, structuring the design of European public services and highlighting why and where interoperability is necessary;
- four levels of interoperability: legal, organisational, semantic and technical;
- the concept of interoperability agreements, based on standards and open platforms.

Finally, the EIF stresses the importance of interoperability governance and the need for coordination across administrative levels. The EIF introduces a conceptual model for developing European public services. It presents a building block approach to constructing them, allowing service components to be interconnected, and promoting the reuse of information, concepts, patterns, solutions, and specifications in Member States and at European level.
Together, the EIS (European Interoperability Strategy) and the EIF are the basis for future activities intended to improve interoperability for delivering European public services. A cross-border interoperability strategy and associated framework have never been attempted on such a scale before, so it will be crucial to involve all stakeholders to ensure success.

The EIS and the EIF will be maintained under the **ISA Programme** (Interoperability Solutions for European Public Administrations):

http://ec.europa.eu/isa/

One of the actions in the ISA Programme 2010-2015 is about architecture:
The Interoperability architecture cluster aims to further align cross-border and cross-sector IT infrastructures that are already available. Creating an interoperability architecture cannot be done overnight. It requires solid, long-term planning and communication. Developments need to take into account the impact on all transactions between the systems involved.
The cluster addresses a broad range of activities:

- Agreeing upon common architecture guidelines
- Creating the architecture itself
- Supporting the maintenance of the architecture
- Identifying and developing common building blocks.

Comment:
The EIF provides a set of principles and recommendations that should be considered for EGDI architecture design.

In the final report "European Interoperability Architecture (EIA) – Phase 2 - Final report: Common Vision for an EIA", version 2.0, November 2011, are defined several interoperability agreements related to the four interoperability levels (legal, organisational, semantic and technical). These agreements are about governance, business, data, applications, technology.

A portal (the "joinup portal", http://joinup.ec.europa.eu/) established with the support of the ISA programme, offers a one stop shop for the sharing and re-use of interoperability assets including vocabularies, code-lists or taxonomies and of best practices for public administrations.

A focus is made on semantic interoperability, providing semantic assets (http://joinup.ec.europa.eu/asset/all) which could be good examples for EGDI to address semantic issues.

8.3 INSPIRE

Description:

The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe.

A European Spatial Data Infrastructure will assist in policy-making across boundaries. Therefore the spatial information considered under the directive is extensive and includes a great variety of topical and technical themes.

INSPIRE is based on a number of common principles:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

Comment:

The technical components of the INSPIRE Architecture have already been implemented in OneGeology-Europe, so they are well-known by the project team, and will be taken into account during the technical design. The INSPIRE Architecture is presented in section 4.5.

8.4 SEIS - Shared Environmental Information System

Description:

The Shared Environmental Information System (SEIS) is a collaborative initiative of the European Commission and the European Environment Agency (EEA) to establish together with the Member States an integrated and shared EU-wide environmental information system. This system would tie in better all existing data gathering and information flows related to EU environmental policies and legislation. It will be based on technologies such as the internet and satellite systems and thus make environmental information more readily available and easier to understand to policy makers and the public.

The underlying aim of SEIS is also to move away from paper-based reporting to a system where information is managed as close as possible to its source and made available to users in an open and transparent way.

According to the SEIS concept, environmentally-related data and information will be stored in electronic databases throughout the European Union. These databases would be interconnected virtually and be compatible with each other. The proposed SEIS is a decentralised but integrated web-enabled information system based on a network of public information providers sharing environmental data and information. It will be built upon existing e-infrastructure, systems and services in Member States and EU institutions.

http://ec.europa.eu/environment/seis/index.htm

[For EGDI technical design, the lessons learned by SEIS have been registered in the following sections]

Learning from experience

Considering lessons learned from past eEnvironment services and activities, could support the identification of a number of patterns on issues to be handled for the successful provision and enhancement of these services. These issues relate to well-known service characteristics, such as the presentation of information derived from significant amounts of data, the collaboration needed for the provision of services, technical decisions etc and issues related to project management and financing.

Ten important aspects are outlined below, which should be considered when designing new or updated eEnvironment services. These issues are presented in no particular order.

mainly because every service may be affected by different factors, depending on many parameters.

eEnvironment services specific issues

Standardisation

Most eEnvironment services summarise vast collections of data, in a form suitable for presentation to their users. Understandably, eEnvironment services require significant

amounts of data originating from a variety of sources. In principle, an eEnvironment service must be designed to facilitate the incorporation of new data sources. This could be an in-feasible task if it has not been considered during the design of the service.
Standardisation should be considered in different levels. The standardisation of metadata is an important step towards the definition of information in any given context (geographical, historical, subjective, etc.). In addition, interoperability, i.e., the standardisation of interfaces, is important when independently created systems need to exchange information for the provision of aggregate services.
An important consideration is the meaning behind any piece of data stored or transmitted. For example, there is a common understanding of what a VAT number is, but every country has its own legislation defining its exact usage. In addition, different systems may store or transmit the same information using different and possibly incompatible representations. Substantial effort for the establishment of a commonly agreed communication method may be required in such cases.
A final consideration relates to terminology. The environment is one of the oldest subjects in every language and many words exist that describe environmental subjects. As a result, significant incompatibilities among languages have developed in environmental terminology.

<u>Collaboration</u>
The natural environment is continuous and is not affected by artificial state borderlines. The environmental preservation requires collaboration among partners from different organisations (possibly from different countries) for the provision of eEnvironment services. However, the collaboration among parties of different organisations can be difficult, due to the diverse priorities, workflows, and policies that could come in the way.
Commonly, there is a necessity of collaboration between partners in a multi-national eEnvironment project. In this case, a contract defines the roles and responsibilities of involved parties, but differences in position could create friction, which may impede on the overall collaboration.
Finally, the collaboration with users must also be taken into account. Such collaboration may not be easy to accomplish, but it is very important as it can provide feedback at early stages of service design and support the delivery of better services.

<u>User friendliness and accessibility</u>
In the design phase of an eEnvironment service, the user's perspective must remain at the centre of attention. The fact that the environment is of general interest regardless of the users' social, educational, professional background, or physical and mental capabilities requires that eEnvironment services are accessible, user friendly and can be used by non-specialists. Neither extensive computer skills, nor environmental expertise should be required in order to request and receive an eEnvironment service. For example, a French project on soil (GIS Sol – National Interest Group on the Soils) considers that computerized soil maps remain difficult to be used by non-specialist users.

<u>Availability of sources</u>
eEnvironment services more often than not require access to large amounts of information. Environmental information is decentralised and comes from a variety of origins: metering

stations, regional or national statistics, citizen reviews to name a few. The availability of sufficient information is an issue that cannot be overlooked, since it can result in a service lacking content.

A large amount of environmental information originates from public sources. In the European Union, the PSI Directive facilitates access to this information. However, there are certain cases where this information may be inadequate or it may need crosschecking and validation. Access to additional sources may be necessary in this case.

Automation of reporting

Linked to the availability of data sources is the mechanism that facilitates the dissemination of environmental information throughout the network of environment agencies at all levels. The automation of reporting could be the only way to makes this process possible, considering the significant amount of data collected. The recommendation to build eEnvironment services close to the information sources is partially based on the facilitation of the reporting process.

Decentralised integration of services

Environmental data such as measurements have characteristics of both time and place. The design of eEnvironment services can benefit from web technologies that facilitate the consolidation and integration of information from different points in time or places on the user's computer. This reduces the complexity of the service, as it does not have to take into account every possible consolidation scenario. For example, in "NatureFrance, Portal of the SINP - Information System on Nature and Landscapes" an integrated WMS (Web Map Service) viewer was used to display geographical information from different servers using simple overlapping, on the client's side.

General considerations

Commitment

For any multi-party cooperation, the terms commitment and motivation are commonplace. The need for commitment is evident, especially in services such as eEnvironment services, which require the collaboration of different parties adequately performing. This commitment may stem from contractual obligation, political will, or financial regulations. Regardless of the motivation, overall success mainly depends on the commitment received as shown from lessons learnt of the French Environmental Portal.

Financing

A financial plan that can secure all involved resources is necessary for any service to succeed. This statement does not suggest that significant financial resources are always required; however, the necessary financial resources should be available.

Managerial issues

Every multi-party project requires sound management. This holds true also for projects related to the provision of eEnvironment services. Such projects deal with parties from different organisations possibly coming from different countries, which have to collaborate for the provision of services. Clear responsibilities, and generally agreed upon procedures and workflows are necessary. Depending on the nature of the tasks involved, the auditing of project activities may range from desirable to compulsory.

Technical issues

As with every service, the technical aspects of the informatics' solution that will implement it must be considered well beforehand. Several viewpoints exist with respect to the service provision paradigm (e.g., client-server, web services), hardware and software assets to be employed, sustainability, service evolution, etc. In a fast-moving information technology world, it is not always easy to select the best mix of available technologies. However, this is important since it can affect the service cost, efficiency, effectiveness, and ultimately, success.

Conclusion

As with any complex project bringing together two different professional worlds, the design and implementation of eEnvironment services should consider several aspects that substantially define the overall project success. Several aspects such as metadata, format, and interface standardisation; collaboration in different levels; availability of source data; user friendliness, and accessibility; and technical features such as reporting automation and decentralised integration – are only a few that should be considered. In addition, other elements to consider would be the motivation of involved parties, financial resources, and other managerial and technical issues.

http://ec.europa.eu/environment/seis/leasons_learned.htm

SEIS provides also the experience about "Facilitating environmental reporting":

Successful decision-making is primarily based on timely and accurate information. As far as the natural environment is concerned, the collection of such information is not a trivial task. This is mainly due to the significant volume of raw data needed in order to cover regularly the many different aspects of the environment in large geographical areas. In this direction, ICT comes to the aid as it can provide the means for quick and reliable communication and data transfer.

Focusing on Europe, three factors make the environment information processing a complex task:

- The vast and increasing number of data collection points and volume of data collected;
- Decentralised governance;
- The correlations that exist between information on related subjects, in different geographical locations, and in different points in time.

These factors result in the need for a structured approach to the information reporting mechanism that must cope with issues at different levels:

- Governance (at the European, national, and regional level)
- Organisational
- Semantic

- Technological

EU Member States all have their own national governments and public administration structures. This complicates the governance and organisational aspects of European-wide environmental reporting. Despite the unanimously declared common interest in the environment preservation, moving in concrete steps is not an easy task.

At the organisational level, it is necessary to take into account the reuse and evolution of existing infrastructure towards a common reporting scheme. It is important to consider extensibility as one of the main concerns of eEnvironment reporting design, as it is more than probable that the needs for information processing will increase.

Additional complexity is inevitable when moving to reporting mechanisms that cover a wider spectrum of environmental parameters, geographical area covered, resolution, etc. However, it is imperative that user interfaces follow the opposite trend: they should become more user-friendly, easier to function, and they should not require extensive IT expertise to operate.

The processing of raw data for the production of statistical and other types of information needs to follow strict methodology requirements. It is therefore important that data and processes that do not follow these requirements can be identified and possibly corrected, even with significant delay. In turn, this requires that all information gathered can be audited and clearly identified parties are associated to them for accountability purposes. It is necessary to maintain the connection between the reporting parties and the data they report so that corrective measures can be taken, when necessary. These requirements evidently dictate the need for named access to the reporting system and the corresponding access control mechanisms.

An important aspect of a common reporting mechanism is eventually to maintain a single definition of terms in use. Previous attempts at such large scale indicate that a viable option is to identify a common data definition that can be accommodated in a common ontology. Subsequently, translation mechanisms can be built in order to allow the conversion between local ontologies to the common one and vice versa. This is not an easy task, however, examples like the current PLOTEUS portal of the European Commission show that it is feasible.

A different aspect is the legal framework under which the reporting mechanism would operate. This becomes a significant concern especially in the cross-border dimension. A pilot operation could be used as a a means to identify the need for specific legal arrangements for cross-border 1 reporting. Such pilot operation should preferably be based on examples of existing reporting obligations that will yield concrete examples of what is usable in real circumstances.

A reporting system for environmental data must act on different administrative levels. It should integrate the access to a large amount of heterogeneous and geographically distributed information of many providers from different administrative levels.

Organisationally, it is necessary that any reporting system is the result of cooperation between environmental agents at different levels. Generally, such collaboration schemes usually include national and regional environment agencies that can interact with each other at the regional, national and international level without significant overhead. It is important that the coordination be made on a national level and that national exchange nodes

are under the control of the Member State. Consequently, and in order to establish a sustainable long-term project, adequate stable funding should be arranged.
Scalability concerns mandate that locally produced information is maintained locally and that it becomes accessible globally. National systems that operate in accordance to this principle could act as a model for a pan-European shared environmental information system (SEIS) not only in its technical but also in the organisational aspects.

Technically, a pan-European eEnvironment reporting system must be distributed. It must leave the information at the providers but offer a single access point to all information to its users. It should also support all existing and planned interfaces within the scope of environmental or geographical information and it should comply with all requirements of the INSPIRE directive and with ISO standards related to geographical information. Distributed architectures have been discussed in SEIS for more than two years and they seem to be the only viable approach. The Member States have to provide the interface and deliver the functionality commonly agreed. This is a lesson coming also from the American CDX portal.

As many functional requirements exist for a pan-European environmental reporting mechanism, it is necessary that a modular approach be adopted so that every involved party can implement a well-defined interface for each module necessary. The following list presents an outline of possible functional modules that need to be considered:

- User management module: Because of the need for accountable reporting, it is necessary that named access to the system be provided.
- Communications module: The multitude of stakeholders, both at the reporting side and at the environmental indicator evaluation side requires a well-defined interface. Different communication principles (off-line, on-line, messaging, real-time, etc.) may need to be taken into account. Both push and pull modes of data transfer need to be supported.
- Content management module: eEnvironment information may exist in any form and the reporting mechanism may not apply restrictions in data types and file formats. For this reason, it is necessary that a content management system be in place to facilitate the identification of information from a variety of sources and data formats. The content management module may provide a wide range of functions that range from a fundamental content repository to enhanced versioning, notification, searching, metadata handling, etc. It is also necessary that access control mechanisms assure that only the right people can have the right access to the right information.
- Auditing module: This module must be able to act independently from the normal reporting procedures. It must be able to provide historical data on reported information and on actions performed, such as downloads, uploads, modifications, deletions, etc.
- Interface module: The number of involved stakeholders mandates that a commonly agreed interface method so that the reporting system can easily integrate with existing or new external systems. The interface module should also take into account human interface at environmental portals, or other methods of access.
- Security module: As per normal practice, the provision of security functionality must be considered from the early design phase. The security module will provide identification, authentication, access control, and any other necessary functionality deemed necessary, e.g., data encryption/decryption, and secure communication in an interoperable manner.

- Additional functionality: Other modules may be considered necessary so that additional functionality may be provided in a commonly agreed form. Such functionality may include the provision of supporting documentation (e.g., help, news, tutorials, etc.) multilingual interfaces, accessibility requirements fulfilment, data format conversions, etc.

To summarise, the concept of an eEnvironment reporting mechanism for Europe is a non-trivial exercise that requires the collaboration of many stakeholders with different function but the same vision: the protection of the natural environment. This exercise needs to consider different levels of interaction that range from governance to the purely technical. Valuable support can be found in good practice and the tools that the information and communication technology can offer. The realisation of such a system is a long-term gradual process.

http://ec.europa.eu/environment/seis/ereporting.htm

Comment:
The lessons learned during the SEIS implementation are of interest for the technical design phase (a few projects describe this point).

SEIS is also an example of “reporting infrastructure”: If, according to the user needs, EGDI has to deal with reporting indicators, the SEIS architecture will provide the technical design phase with interesting recommendations.
Two general recommendations:

- The distributed architectures, discussed in SEIS for more than two years, seem to be the only viable approach.
- The modular approach, specifying user management, communications, content management, auditing, interface, security modules should be also considered for EGDI..

9 Standards

To implement an open and interoperable infrastructure, most of projects and framework documents request to use standards for all parts of the infrastructure. In the SDI domain, ISO, CEN and OGC are the key standards providers on top of very technical standards produced by W3C, OASIS, IEEE.

9.1 ISO - RM-ODP: Reference Model

RM-ODP: Reference Model for Open Distributed Processing.

The objective of standardization is the development of standards that allow the benefits of distributing information processing services to be realized in an environment of heterogeneous IT resources and multiple organizational domains. These standards address constraints on system specification and the provision of a system infrastructure that accommodate difficulties inherent in the design and implementation of distributed systems.
RM-ODP provides the coordination framework for these ODP standards. It creates an architecture within which support of distribution, interworking and portability can be integrated.
ODP standardization has four fundamental elements:

- A guide for architects to specify distributed software systems. It is based on object modelling, which is relevant to the practices of systems architects;
- The specification of a system in terms of separate but interrelated viewpoints;
- The definition of a system infrastructure providing distribution transparencies for system applications;
- A framework for assessing system conformance.

RM-ODP allows describing complex distributed systems and defines a framework comprising five viewpoints: Enterprise, Information, Computation, Engineering and Technology.

RM-ODP viewpoints :

Viewpoint	**Definition of RM-ODP Viewpoint (ISO/IEC 10746-1)**
Enterprise viewpoint	A viewpoint of an ODP system and its environment that focuses on the purpose, scope and policies of that system.
Computational viewpoint	A viewpoint on an ODP system and its environment that enables distribution through functional decomposition of the system into objects which interact at interfaces.
Information viewpoint	A viewpoint on an ODP system and its environment that focuses on the semantics of information and information processing.

Engineering viewpoint	A viewpoint on an ODP system and its environment that focuses on the mechanisms and functions required to support distributed interaction between objects in the system.
Technology viewpoint	A viewpoint on an ODP system and its environment that focuses on the choice of technology in that system.

Comments:
The recommendation for EGDI is to use this standard, as GEOSS AIP and other projects, to design the architecture.

9.2 OGC/ISO – Geographic Information

In the SDI domain, most of standards are established by OGC, ISO/TC 211, CEN/TC 287. Many of them are used to implement services, metadata, data models, …in numerous projects analysed in this document, but also for international initiatives as GEOSS, OneGeology or directives as INSPIRE.

The list below shows the topics addressed by these standards:

EN ISO 19101, Geographic information — Reference model.
ISO/TS 19101-2 Reference model - Part 2: Imagery
ISO/TS 19103, Geographic information — Conceptual schema language
ISO/TS 19104, Geographic information — Terminology
EN ISO 19105, Geographic information — Conformance and testing
EN ISO 19106, Geographic information — Profiles
EN ISO 19107, Geographic information — Spatial schema
EN ISO 19108, Geographic information — Temporal schema
EN ISO 19109, Geographic information — Rules for application schema
EN ISO 19110, Geographic information — Feature cataloguing methodology
EN ISO 19111, Geographic information — Spatial referencing by co-ordinates
EN ISO 19112, Geographic information — Spatial referencing by geographic identifiers
EN ISO 19113, Geographic information — Quality principles
EN ISO 19114, Geographic information — Quality evaluation procedures
EN ISO 19115, Geographic information — Metadata
ISO 19115-2 Metadata - Part 2: Extensions for imagery and gridded data
EN ISO 19116, Geographic information — Positioning services
EN ISO 19117, - Geographic information — Portrayal
EN ISO 19118, Geographic information — Encoding
EN ISO 19119, Geographic information — Services
ISO/TR 19121 Geographic information - Imagery and gridded data
EN ISO 19123, Geographic information — Schema for coverage geometry and functions
EN ISO 19125-1, Geographic information — Simple feature access — Common architecture
EN ISO 19125-2, Geographic information — Simple feature access — SQL option
EN ISO 19126 Geographic information -- Feature concept dictionaries and registers.

ISO/TS 19127 Geographic information — Geodetic codes and parameters

EN ISO 19128, Geographic information — Web map server interface

ISO/TS 19129, Geographic information – Imagery, gridded data framework

EN ISO 19131, Geographic information — Data product specification

EN ISO 19132, Geographic information — Location based services possible standards

EN ISO 19133, Geographic information — Location based services tracking and navigation

EN ISO 19134, Geographic information —Location based services — Multimodal routing and navigation

EN ISO 19135, Geographic information — Procedures for registration of geographic information items

EN ISO 19136, Geographic information — Geography Markup Language (GML)

EN ISO 19137, Geographic information — Core profile of the spatial schema

ISO/TS 19138, Geographic information — Data Quality Measures

ISO/TS 19139, Geographic information — Metadata — Implementation specification

EN ISO 19141, Geographic information — Schema for moving features

EN ISO 19142, Geographic information — Web feature service

EN ISO 19143, Geographic information — Filter encoding

EN ISO 19144-1, Geographic information - Classification systems - Part 1: Classification system structure

Comments:
The recommendation for EGDI is to use these standards when relevant for the architecture design.

9.3 CEN- Geographic Information - Spatial Data Infrastructures

With RM-ODP, ISO delivers a standard for high level design, and with the 19000 series, OGC, ISO/TC 211 and CEN/TC 287 provides standards for the infrastructure components. The technical report CEN/TR 15449:2011 "Geographic information – Standards, specifications, technical reports and guidelines, required to implement Spatial Data Infrastructures" focuses on technical aspects to implement neutral-technological infrastructure for geospatial data and services, based upon standards and specifications.

It considers an SDI as a collaborative framework of disparate information systems that contain resources that stakeholders desire to share.

Considering the complexity of the subject and the need to capture and formalize different conceptual and modelling views, TR 15449 is comprised of multiple parts:

- Part 1: Reference model: this provides a general context model for the other parts, applying general IT architecture standards;
- Part 2: Best Practice: This provides best practices guidance for implementing SDI;.
- Part 3: Data centric view: This addresses concerns related to the data, which includes application schemas and metadata.
- Part 4: Service centric view: This includes the taxonomy of services, concepts of interoperability, service architecture, service catalogue, and the underlying IT standards.

Comments:

This report is of great interest for the EGDI Architecture design, as it is a summary of the description of many useful SDI components and recommendations from best practices.