FP7-INFRASTRUCTURES-2012-1

Grant Agreement no. 312845

Scoping Study for a pan-European Geological Data Infrastructure

D 4.2

Report on data models, vocabularies, services, and portal specifications for the EGDI

Deliverable number	*D4.2*
Dissemination level	*Public*
Delivery date	*28 February 2014*
Status	*Final*
Author	*Jean-Jacques Serrano, BRGM*

EGDI-Scope Project, WP4

Report on data models, vocabularies, services, and portal specifications for the EGDI

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 purpose of this report is to complete the previous deliverable (D4.1, an inventory of existing infrastructures and related components) with an inventory on metadata, data models, and services of interest for EGDI. This report is an input for the next deliverable D4.3 about architecture needs which will take also into account requirements from WP2 and WP3.

This report is a report on data models, vocabularies, services, and portal specifications for the EGDI. It identifies existing standards for data models developed through international or European initiatives (INSPIRE core and extensions), and their appropriateness for EGDI. It includes the standards for metadata and for semantic interoperability (vocabularies or code-lists, and related web services).
It identifies required services for EGDI, including: access control and security, data publishing, discovery, conformity testing, view (including on 3D models), data access, registry management and access, monitoring of the infrastructure services.
It includes also some specifications of a multilingual EGDI portal giving access to all datasets, including 3D data and models.

2 Metadata about datasets and services

Metadata are a set of elements able to describe datasets and services, so that a user can discover, evaluate and use them.
EGDI resources should be visible in other infrastructures, as EPOS for example, and the first option to be connected is to have visible metadata (and discovery services). The recommendation for EGDI is to use the INSPIRE metadata regulation, but other metadata standards must be known (for example EPOS uses CERIF). That does not mean EGDI should deliver metadata according to several standards but solution will have to be identified to make EGDI visible into other systems (as a mapping of CERIF metadata for EPOS with INSPIRE/ISO 19115 metadata – mapping on-going)
Various metadata standards are available, which could be of interest for EGDI resources:

- INSPIRE metadata, based on the ISO standards 19115 & 19119
- Dublin Core
- CERIF (Common European Research Information Format)
- DCAT-AP (Application Profile of Data Catalogue vocabulary)
- CKAN (Comprehensive Knowledge Archive Network)

2.1 INSPIRE Metadata

The INSPIRE metadata elements are based on the ISO standards 19115 & 19119, and currently address the “discovery level”. Metadata elements to address the “evaluate” and “use” levels are not yet in the scope of the current INSPIRE regulation. But several ISO elements could be used for these levels.

List of INSPIRE Metadata elements, organised in 10 sections:

1. IDENTIFICATION
1.1. Resource title
1.2. Resource abstract
1.3. Resource type
1.4. Resource locator
1.5. Unique resource identifier
1.6. Coupled resource
1.7. Resource language

2. CLASSIFICATION OF SPATIAL DATA AND SERVICES
2.1. Topic category
2.2. Spatial data service type

3. KEYWORD
3.1. Keyword value
3.2. Originating controlled vocabulary

4. GEOGRAPHIC LOCATION
4.1. Geographic bounding box

5. TEMPORAL REFERENCE
5.1. Temporal extent
5.2. Date of publication
5.3. Date of last revision
5.4. Date of creation

6. QUALITY AND VALIDITY
6.1. Lineage
6.2. Spatial resolution

7. CONFORMITY
7.1. Specification
7.2. Degree

8. CONSTRAINT RELATED TO ACCESS AND USE
8.1. Conditions applying to access and use
8.2. Limitations on public access

9. ORGANISATIONS RESPONSIBLE
9.1. Responsible party
9.2. Responsible party role

10. METADATA ON METADATA
10.1. Metadata point of contact
10.2. Metadata date
10.3. Metadata language

The two elements "1.6. Coupled resource" and "2.2. Spatial data service type" are not applicable for datasets but only for services.

More information about INSPIRE metadata :
http://inspire.jrc.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf

2.2 Dublin Core

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. These elements are broad and generic, usable for describing a wide range of resources.

The fifteen elements "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metada-

ta Initiative (DCMI). The full set of vocabularies, DCMI Metadata Terms, also includes sets of resource classes (including the DCMI Type Vocabulary), vocabulary encoding schemes, and syntax encoding schemes. The terms in DCMI vocabularies are intended to be used in combination with terms from other, compatible vocabularies in the context of application profiles and on the basis of the DCMI Abstract Model.

List of Dublin Core metadata elements:

Contributor	An entity responsible for making contributions to the resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Creator	An entity primarily responsible for making the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Description	An account of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Language	A language of the resource.
Publisher	An entity responsible for making the resource available.
Relation	A related resource.
Rights	Information about rights held in and over the resource.
Source	A related resource from which the described resource is derived.
Subject	The topic of the resource.
Title	A name given to the resource.
Type	The nature or genre of the resource.

More information about Dublin Core metadata :
http://dublincore.org/documents/dces/

2.3 CERIF

CERIF: Common European Research Information Format
The CERIF data model allows for a metadata representation of research entities, their activities / interconnections (research) and their output (results) as well as high flexibility with formal relationships, enables quality maintenance, archiving, access and interchange of research information and supports knowledge transfer to decision makers, for research evaluation, research managers, strategists, researchers, editors and the general public.
CERIF is implemented in EPOS.
A CRIS (Current Research Information System) can be implemented using a subset or superset of the full CERIF model for projects, people, organisations, publications, patents, products, services and facilities (equipment in particular) with role-based, temporally-bound relationships.

Objects or entities managed by CERIF:

- Researchers
- Organisations (Research-performing, Funding)
- Funding Programmes, Calls
- Projects (Proposed, Ongoing, Completed)
- Publications, Patents, Data, Products
- Facilities, Equipment, Services
- Addresses, Geographic Bindings, Languages

All these entities are connected by relationships providing semantics including roles and time. CERIF is extensible to address richer uses.

More information about CERIF:
http://www.eurocris.org/Index.php?page=CERIFintroduction&t=1

2.4 DCAT Application profile for data portals in Europe

DCAT-AP: Application Profile of Data Catalogue vocabulary

DCAT is an RDF (Resource Description Framework) vocabulary designed to facilitate interoperability between data catalogues published on the Web.

The purpose of DCAT-AP is to define a metadata profile based on the W3C Data Catalogue vocabulary (DCAT) that can be used as a common interchange format for data portals of the EU and of EU Member States.

DCAT-AP metadata elements:

DCAT-AP Class	Definition / Properties
Agent	An entity that is associated with Catalogues and/or Datasets Properties: Name.
Category	A subject of a Dataset. Properties: Preferred label
Category scheme	A concept collection (e.g. controlled vocabulary) in which the Category is defined Properties: Title
Catalogue	A catalogue or repository that hosts the Datasets being described. Properties: Dataset; Title; Description; Licence; Publisher Home page; language; licence; release date; themes; update / modification date; record; rights; spatial / geographic
Catalogue Record	A description of a Dataset's entry in the Catalogue Properties: primary topic; update / modification date; listing date; change type Description; Title
Dataset	A conceptual entity that represents the information published. Properties: Title; Description; Publisher Contact point; Dataset distribution; keyword / tag; theme / category Conforms to; frequency; identifier; landing page; language; other identifier; release date; spatial / geographic coverage; temporal coverage; update / modification date; version; version notes
Distribution	A physical embodiment of the Dataset in a particular format. Properties: Access URL; Licence; Description; Format Byte size; Download URL; Media type; Release date; Rights; Status; Title; update / modification date
Document	A textual resource intended for human consumption that contains information, e.g. a web page about a Dataset
Frequency	A rate at which something recurs, e.g. the publication of a Dataset.
Identifier	An identifier in a particular context, consisting of the string that is the identifier; an optional identifier for the identifier scheme; an optional identifier for the version of the identifier scheme; an optional identifier for the agency that manages the identifier scheme
Licence document	A legal document giving official permission to do something with a resource. Properties: Licence type

DCAT-AP Class	Definition / Properties
Literal	A literal value such as a string or integer; Literals may be typed, e.g. as a date according to xsd:date. Literals that contain human-readable text have an optional language tag Properties:
Linguistic system	A system of signs, symbols, sounds, gestures, or rules used in communication, e.g. a language
Location	A spatial region or named place. It can be represented using a controlled vocabulary or with geographic coordinates. In the latter case, the use of the Core Location Vocabulary is recommended.
Media type or extent	A media type or extent, e.g. the format of a computer file
Period of time	An interval of time that is named or defined by its start and end dates. Properties: start date / time; end date / time
Publisher type	A type of organisation that acts as a publisher
Rights statement	A statement about the intellectual property rights (IPR) held in or over a resource, a legal document giving official permission to do something with a resource, or a statement about access rights.
Standard	A standard or other specification to which a Dataset conforms
Status	An indication of the maturity of a Distribution.
VCard	A description following the vCard specification, e.g. to provide telephone number and e-mail address for a contact point.
Resource	Anything described by RDF.

The scope of DCAT-AP is broader than INSPIRE metadata, since it covers any kind of EU data, even those not covered by the INSPIRE Directive. But an on-going proposal is to improve the compatibility between INSPIRE metadata and DCAT-AP.

More information on DCAT:
https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final

2.5 CKAN

CKAN (Comprehensive Knowledge Archive Network) developed by the Open Knowledge Foundation, is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data. CKAN is aimed at

data publishers (national and regional governments, companies and organizations) wanting to make their data open and available.

Each dataset is described by a set of elements: Title, Unique identifier, Groups (similar to categories), Description, Revision history, Licence, Tags, Formats.

CKAN is one of the components of the Open Data Portal containing data from the European Commission (https://open-data.europa.eu/en/data/)

2.6 Comments for EGDI

INSPIRE Metadata are mandatory for Member States to describe datasets and services related to geosciences data themes (geology, mineral resources, natural hazards …).
They have been already implemented in some projects as OneGeology-Europe or ProMine by the Geological Surveys.
So the recommendation for EGDI is clearly to use the INSPIRE metadata.

The connection of INSPIRE metadata with EPOS/CERIF metadata is currently being defined, so if EGDI metadata are described according to the INSPIRE regulation on metadata it will be possible for EPOS users to discover geological datasets.
Datasets described with INSPIRE metadata should be visible in the DCAT-AP as there is a mapping between INSPIRE metadata elements and DCAT-AP elements (even if some issues are still to be solved).
The INSPIRE metadata are a set of elements to address the first level (discovery level) of metadata. The other levels (evaluate and use) could be addressed by EGDI using other metadata elements defined in the ISO 19115/19119 standards.

The connection of EGDI metadata to the CKAN registry should be taken into account for datasets delivered to EC. CKAN can already import metadata from an OGC CS-W server.

3 Data models

3.1 Datasets for EGDI

The EGDI-Scope WP3 has selected as a priority the following thematic areas:

- Geology
- Mineral Resources
- Water Resources
- Geohazards: flooding, earthquakes, landslides, subsidence
- Soil

3.2 Data models

The data models presented in this document are for data to be delivered at the European level to provide European products, and not to address national or local issues that could request more detailed information.
For the datasets selected by WP3, standards data models exist, defined by the INSPIRE Data Specifications.
These European data specifications contain a core data model describing the main classes and properties (often mandatory) and suggest to extend the data model to address more use cases than the ones used to define the core model. The main goal of INSPIRE is to build a **spatial** infrastructure, the thematic aspects are covered by specific directives (as the Water Framework Directive or the Mining Waste Directive) requesting more properties.

For the 5 thematic areas selected for EGDI data specifications (including a data model) have been specified in the INSPIRE Directive. For some of them an international standard data model also exist, often used to specify the INSPIRE model.

This section does not describe in details all the data models as they are presented in the INSPIRE documents (Data Specifications / Technical Guidelines)

3.3 Data modelling principles and common rules

To enable the interoperability and, where practicable, harmonisation of spatial data sets and services within Europe, INSPIRE has defined common rules for data modelling in the **Generic Conceptual Model** document.
This document defines rules for application schemas, coordinate referencing and units model, identifier management, multi-lingual text and cultural adaptability, object referencing modelling, multiple representations (levels of detail) and consistency, and more.

Common or shared spatial object types relevant in multiple themes are identified and modelled in a consistent manner. Spatial characteristics of a spatial object are represented by vector geometries, coverage functions.

The Generic Conceptual Model provides requirements and recommendations regarding the following aspects:

- INSPIRE application schemas
- spatial and temporal representations of spatial objects across different levels of detail
- spatial and temporal relationships between spatial objects
- unique object identifiers
- constraints
- reference to common spatial and temporal reference systems
- controlled vocabularies
- support for multilingual aspects

More information about the Generic Conceptual model:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf

Related to the Generic Conceptual Document, INSPIRE has defined guidelines for the encoding of spatial data for the purpose of data interchange between systems in INSPIRE (access to data via services which includes but is not limited to a download of a complete spatial data set).
More information about the guidelines for the encoding of spatial data:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.7_v3.3rc3.pdf

Other common rules

As various INSPIRE domains use the same type of objects/classes some common rules are specified for:

- Coverages (gridded data): Application Schemas for coverage types
  http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.10.2_CoverageTypes_v1.0rc3.pdf
- Networks (hydrography, transports, …): Generic application schema for networks
  http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.10.1_GenericNetworkModel_v1.0rc3.pdf
- Use of Observations & Measurements and Sensor Web Enablement:
  http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.9_O&M_Guidelines_v2.0rc3.pdf

3.4 Comments for EGDI about common rules

All the requirements specified in the Generic Conceptual Model have been taken into account for the data specifications of the thematic domains of EGDI. But for the future of EGDI addressing more classes and properties than those specified in the data specifications, or addressing other thematic areas, it is recommended to define the new data models or

the data model extensions according to the Generic Conceptual Model recommendations and other common rules (about coverages, network, observations & measurements).

3.5 Data model for Geology

The core data model contains the main types of Geologic Features (Geologic Units, Geologic Structures, and Geomorphologic Features). The geometry of these features is described in Mapped Features and can be included in geological maps and profiles in the form of points, lines and polygons. The data model also enables a description of the lithological/stratigraphical characteristics of borehole logs, thematic maps, geophysical surveys and measurements, and features related to hydrogeology (aquifers and groundwater bodies).
Main classes of the Geology data model:

The abstract **Geologic Feature** class represents a conceptual geological feature composed by three sub-types: Geologic Unit, Geologic Structure and Geomorphologic Feature. A **Mapped Feature** provides a spatial representation of a Geologic Feature.

- **Geologic Unit** represents a body of material in the Earth whose complete and precise extent is inferred to exist. Spatial properties are on available through association with a Mapped Feature. The composition association from Geologic Unit to **Composition Part** allows the lithological description of the Geologic Unit. The composition of a Geologic Unit can be made up of several Composition Parts, for example where there are lithologically distinct components interbedded.
- **Geologic Structure** is defined as a configuration of matter in the Earth based on describable inhomogeneity, pattern, or fracture in an Earth Material. The identity of a Geologic Structure is independent of the material that is the substrate for the structure. The two types of Geologic Structure in the data model are Shear Displacement Structure and Fold.
- The abstract **Geomorphologic Feature** class is a point, linear or areal landform or landscape. It is a natural or an anthropogenic surface feature and may be erosional, depositional or both. Geomorphologic Feature has two subtypes: **Natural Geomorphologic Feature and Anthropogenic Geomorphologic Feature**.

A **Geologic Event** is defined as an identifiable event during which one or more geological processes act to modify geological entities. Geological age is modelled using Geologic Event – the age of some geological event occurring. A Geologic Event should have a specified geologic age and process, and may have a specified environment.

The **geologic History** association from Geologic Feature to Geologic Event describes a sequence of one or more Geologic Events which together describe the age or geologic history of the Geologic Feature. Commonly Geologic Features will have a geologic History

comprising only one Geologic Event, which represents the formation of the Geologic Feature.

A **Geologic Collection** is a named or identifiable group of geological or geophysical objects. Geologic objects are commonly grouped into collections such as geological maps, thematic maps, groups of geophysical measurements or models of the same type etc., which are familiar to many user communities. The Geologic Collection class allows the delivery of a package of objects that go to make up one of these familiar collections.

Borehole is a generalized class for any narrow shaft drilled in the ground, at any angle. A borehole log is a collection of Mapped Intervals, each off which can be specified by a Geologic Unit and have an age. This allows the description of lithological or stratigraphical borehole logs. A Mapped Interval is a special kind of Mapped Feature whose shape is a 1-D interval and which uses the spatial reference system (SRS) of the containing borehole.

More information about the data model for Geology (UML diagrams and feature dictionary): http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GE_v3.0rc3.pdf

International Standards

For the Geology domain, the international standard is GeoSciML (www.geosciml.org).It has been the source to specify the INSPIRE data model. This standard is suggested in the INSPIRE Specifications to address more uses cases.
The INSPIRE model is a subset of GeoSciML. The main interest to have this standard behind the INSPIRE model is when more classes/properties are necessary to address “real thematic” use cases: the model for these properties is already done and standardized.

In the petroleum domain a standard, RESQML, provides open, non-proprietary data exchange standards for reservoir characterization, earth and reservoir models. It defines standard about structures (faults, horizons, stratigraphic and structural organisation), wells, 3D grids, and “infrastructure” (metadata, formats, reference systems – coordinates, units of measurements)..

3.6 Data model for Mineral Resources

The main spatial objects / data types of the Mineral Resources domain are Mineral Occurrence, Commodity, Mine and Product (mined material or concentrate), Mining Waste, Exploration activity, and Mining activity.
Two application schemas are provided:

- the core data model: Mineral Resources, related to the main object types and properties: the location of mineral resources (Mines and Mineral Occurrences), the main commodities, and the exploitation type,

- the extension: Mineral Resources Extension, to address more properties, e.g. to able to provide more attributes for describing mineral resources, specially to meet requirements from the Raw Materials Initiative and the Mining Waste Directive.

The Mineral Resources data model:

The **Mining Feature** class represents a conceptual feature and corresponds with a "Mine" or a "Mining Activity", locatable and identifiable features in time and/or space. The Mining Feature Occurrence is an occurrence of a Mining Feature. It carries some properties and the geometry and/or location.

- A **Mine** is an excavation for the extraction of mineral deposits. 'True' mines are underground workings and open-pit workings (also called open-sky mines) generally for the extraction of metallic commodities. The Mine spatial object type also includes open workings generally for the extraction of industrial minerals, commonly referred to as quarries.
- The **Mining Activity**, related to a Mine, describes the process of extracting metallic or non-metallic mineral deposits from the Earth.

The **Earth Resource** identifies the kinds of observable or inferred phenomena required to classify economic and sub-economic earth resources:

- The **Mineral Occurrence** could be a prospect, an occurrence, a mineral deposit, an ore deposit, etc. (but not a lode, a field, a district, or a province)
- The **Commodity** describes the material of economic interest in the Earth Resource
- The **Ore Measure** is an estimated or calculated amount of ore and grade that exist within an Earth Resource, in terms of its resource, reserve and endowment
- The **Mineral Deposit Model** describes the essential attributes of a class of mineral deposits used to classify the Earth Resource

An Earth Resource has an associated Exploration Activity to describe the process leading to the discovery and assessment of the resource.

The Mineral Resources data model uses classes from the INSPIRE Geology Data Model: an Earth Resource is a Geologic Feature, which has geometry (a Mapped Feature) and an age (Geologic Event).

Exploration history is needed for quantitative assessment of possibly existing, yet undiscovered mineral resources of an area. Such information can also help to evaluate the potential of an occurrence (sampling survey; drilling survey).

The Mineral ResourcesExtension data model:

This extension of the core Mineral Resources data model provides more attributes describing mineral resources, especially to meet requirements from the EU Raw Materials Initiative and the Mining Waste Directive.

The following types/attributes describe **Mining Waste** and Mining Waste Measure:

- Waste type
- Material
- Processing type
- Storage type

- Waste measures (density, grade per commodity, volume)
- Environmental Impact

The following types/attributes describe **Products** and **Mined Material**:

- Product
- Grade
- Production
- Recovery
- Material
- Raw Material Role
- Proportion

The following types/attributes describe the **composition** of the Earth Resource with Earth Resource Material:

- Material as Earth Material utilizing the GeoSciML definition.
- Proportion of the material in the earth resource
- Role of the material described (host rock, alteration product, …)

More information about the data model for Mineral Resources (UML diagrams and feature dictionary):
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_MR_v3.0rc3.pdf

International Standard

For the Mineral Resources domain, the international standard is EarthResourceML (http://www.earthresourceml.org) It has been the source to specify the INSPIRE data model. This standard is suggested in the INSPIRE Specifications to address more uses cases.

The INSPIRE model is a subset of EarthResourceML. The main interest to have this standard behind the INSPIRE model is when more classes/properties are necessary to address “real thematic” use cases: the model for these properties is already done and standardized.

The UNFC standard (United Nations Framework Classification for Fossil Energy and Mineral Reserves and Resources) is used for the classification and reporting of fossil energy and mineral reserves and resources.

3.7 Data model for Water and Groundwater

The INSPIRE Hydrogeological data model identifies two basic elements: the **'rock' system** or aquifer system (invariable in time) containing hydrogeological units, classified as aquifers, aquitards and aquicludes and the **‘groundwater’ system** with groundwater bodies (variable in time). Hydrogeological objects (man-made and natural objects such as groundwater wells and springs) interact with these domains of the ‘rock’ system and the

'groundwater' system. The 'rock' system and the 'groundwater' system and the interaction between them create a **hydrogeological system**. The principal aim of the core model is to capture the main classes of these systems and to provide the logical links between them. The 'groundwater' system is created by groundwater flow in aquifers of the 'rock' system, which have the right porosity and permeability to conduct groundwater. The 'groundwater' system has distinct groundwater flow properties and a distinct pressure regime and is confined by permeability, groundwater surface or other barriers in the subsurface.

For provision of detailed measurements on the quality and chemical composition of groundwater and time series measurements of groundwater level within groundwater wells the use of the INSPIRE Environmental Monitoring Facilities specification is recommended.

More information about the data model for Environmental Monitoring Facilities:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EF_v3.0.pdf

Main classes of the Groundwater data model:

The 'Rock' system:

The 'Rock' system has one main class, **Hydrogeological Unit**, with a number of important subclasses. Hydrogeological Unit is a part of the lithosphere with distinctive parameters for water storage and conduction, and is a specialisation of Geologic Unit.
There are four important subclasses of Hydrogeological Unit: Aquifer, Aquitard, Aquiclude and Aquifer System

- An **Aquifer** is a wet underground layer of water-bearing permeable rock or unconsolidated materials (gravel, sand, silt, or clay) from which groundwater can be usefully extracted by a groundwater well.
- An **Aquitard** is a saturated, but poorly permeable bed that impedes groundwater movement and does not yield water freely to wells, but which may transmit appreciable water to or from adjacent aquifers and, where sufficiently thick, may constitute an important groundwater storage unit.
- An **Aquiclude** is a Hydrogeological Unit that due to its low permeability can act as a barrier to groundwater flow and as such often confines aquifers or aquifer systems.
- An **Aquifer System** is a collection of Aquifers and/or Aquitards which together constitute the environment of groundwater - "communicating vessels" that are filled or can be filled with groundwater i.e. a Ground Water Body. An Aquifer System may contain one or more Aquifers, Aquitards and Aquicludes .

The 'Groundwater' system:

The main class of the groundwater system is Ground Water Body.

- A **Ground Water Body** is a distinct volume of groundwater within an aquifer or system of aquifers, which is hydraulically isolated from nearby groundwater bodies. The piezometric State property of a Ground Water Body, which specifies the piezometric state of the groundwater body water table, is modelled in a separate class Piezometric State. Ground Water Body interacts with the 'rock' system through an association with Aquifer System

- **WFD Ground Water Body** is a distinct volume of groundwater within a groundwater flow system, which is used as a reporting or management unit within the Water Framework Directive (WFD).

The **hydrogeological system** is formed by the interaction of the groundwater system and the rock system.

Hydrogeological Object:
Hydrogeological Object is an abstract class for man-made or natural objects where interaction occurs with the hydrogeological system. HydrogeologicalObject has two subclasses HydrogeologicalObjectManMade and HydrogeologicalObjectNatural

- **HydrogeologicalObjectManMade** is an abstract class for a manmade facility, where interaction occurs with the hydrogeological system.
  An Active Well is the only type of HydrogeologicalObjectManMade defined in this application schema. It is an excavation or opening into the ground where the intended use is for location, acquisition, development, or artificial recharge of ground water.
- **HydrogeologicalObjectNatural** for natural objects where interaction (inflow or outflow) occurs with the hydrogeological system.

More information about the data model for Groundwater, included into the Geology theme (UML diagrams and feature dictionary):
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GE_v3.0rc3.pdf
More information about the data model for Water:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_HY_v3.0.1.pdf

International Standard
For the Water and Groundwater domain, the international standard are WaterML 2.0 and GroundwaterML (the previous version of WaterML addressed only the CUASHI needs (CUASHI: Consortium of Universities for the Advancement of Hydrologic Science, www.cuahsi.org):
http://external.opengeospatial.org/twiki_public/HydrologyDWG/WebHome

They have been the source to specify the INSPIRE data model. This standard is suggested in the INSPIRE Specifications to address more uses cases.

3.8 Data model for Soil

The Soil data model covers:

- Soil inventories, providing one-off assessments of soil conditions and/or soil properties at certain locations and at a specific point in time, and allow soil monitoring,

providing a series of assessments showing how soil conditions and/or properties change over time.
- Soil mapping, providing a spatial presentation of the properties linked to the soils, including soil types; typically, soil maps are derived with the help of data available in soil inventories. Also other soil related information derived from soil properties, possibly in combination with non-soil data are within the scope.

And includes:
- soil profiles
- soil sites, soil plots
- soil bodies (delineated areas on the earth's surface determined on the basis of certain soil characteristics)
- soil characteristics (parameters) that change over time (possibly allowing soil monitoring)

Soil profile:
A soil profile is a cross-section of the soil from the surface down to and including the beginning of the fresh material unmodified by pedogenesis, consisting of various more or less horizontally oriented features formed by pedogenic processes called horizons. Any real world soil profile, considered as a whole, can be characterized by a number of properties, such as: its soil type according to a soil classification scheme, its water capacity available for plants, coarse fragment cover, the presence of a water table, etc.
- **Soil horizons** within a profile possess biological, physical and chemical characteristics which differ from the horizons above and beneath and are distinguished principally by texture, colour, structure, chemical composition, and their biomass. The chemical properties may include concentrations in the solid, water and air phases, mobility and soil adsorption capacities.
- A profile can additionally be described with **layers** (instead of being described with horizons) which do not necessarily correspond with pedo-genetically formed horizons. Such layers can be fixed depths intervals of sampling, top- and subsoil, or geogenic layers.

An **Observed Soil Profile** represents a soil profile physically located in a soil plot (or exposed with a boring), described in the field, possibly sampled and analysed in the laboratory. An observed soil profile refers to a real world location (specified by an associated soil plot).
A **Derived Soil Profile** is a soil profile that cannot be located in a soil plot. It corresponds to the spatial extent of a soil type that has been observed in one or several observed soil profiles. The derived soil profile has property values that could be derived (e.g. averaged) from the values of the corresponding properties of one or more observed soil profiles. The derived soil profile can be characterized by the same properties as those of the observed soil profiles, but it is understood that the values for these properties have been derived or determined by expert judgment or calculation.

Soil delineated areas
To delineate spatially an area that is characterized by a set of derived soil profiles, the model introduces the construct of "**Soil Body**" which represents an association of soils that are found together in the area. A soil body can be interpreted as a Soil Mapping Unit (SMU) as recognized in the European Soil Geographical Data Base and other soil databases at national or regional levels.

Soil Site and Soil Plot:
A **soil site** is considered as the surrounding of a soil profile, and/or the larger piece of land that is directly linked to and described by all soil investigations on one or more spots, called soil plots.
A **soil plot** is the location of a specific soil investigation (e.g. of a trial pit).

Soil Theme Coverage:
The specific purpose of the Soil Theme Coverage class is to provide a structure for the interchange of soil thematic maps as continuous phenomena. The geometry is represented as a coverage.

More information about the data model for Soil:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SO_v3.0rc3.pdf

3.9 Data model for Natural hazards

In this data model Natural Risk Zones are defined as areas where natural hazards are co-incident with populated areas and/or areas of particular environmental / cultural or economic value. Risk in this context is defined as: Risk = Hazard x Exposure x Vulnerability of human health, the environmental, cultural and economic assets in the zone considered

Different types of hazard are considered:

- Floods (calculation of flood impact, reporting and flood hazard/risk mapping)
- Risk Management Scenario (an example from a national perspective)
- Landslides (hazard mapping, vulnerability assessment and risk assessment)
- Forest fires (danger, vulnerability and risk mapping)
- Earthquake insurance

The concepts included in the model are abstract and can be specialised using either vector or coverage spatial representation. This is done in order to create a framework which enables exchange of data that are either vector or coverage, considering that any of the spatial objects can be modelled in either way.

There are 4 key spatial object types that are modelled (as vectors and as coverages);

- Hazard area
- Observed event
- Risk zone
- Exposed element

For each of them, 3 spatial object types are created:

- An abstract spatial object type that contains the properties (attributes, or constraints) of the spatial object that are common both to its vector representation and to its coverage representation. These abstract spatial object types have their names prefixed by “Abstract”.
- A vector spatial object type that is generated from the abstract spatial object. It has the properties that are specific to the vector representation, such as the definition of the geometry.
- A coverage spatial object type that is generated both from the abstract spatial object and a generic coverage spatial object type (detailed later in the chapter). It has the properties that are specific to coverage representation, such as the definition of the domain and the definition of the range. These coverage spatial object types have their names suffixed by “Coverage”.

A **Hazard area** has a perimeter delineated according to a specific method of determination (by modelling or by indirect determination), a type of hazard and a period of validity. This area could be linked to Observed events. Other properties of the hazard area are the likelihood of occurrence and the magnitude

Exposed element refers to the spatial representation of people, property, systems, or other elements present in hazard zones that are thereby subject to potential losses. The assessment or calculation of vulnerability may be conducted over those spatial objects.

A **risk zone** is defined as the spatial extent of a combination of the consequences of an event (hazard) and the associated probability/likelihood of its occurrence. The property “source of risk” refers to the type of hazard that engenders the risk, and the “level of risk” is an assessment of the combination of the consequences of an event (hazard) and the associated probability/likelihood of the occurrence of the event

An **observed event** refers to the spatial representation of a natural phenomenon relevant to the study of natural hazards which occurred, or is currently occurring, and which have been observed. The type of hazard, the name of event, the level of intensity is some properties of the observed event.

More information about the data model for Natural hazards:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_NZ_v3.0rc3.pdf

International Standard

There are no adopted international standards for the Geohazard domain, even if for some there are some proposals as QuakeML (https://quake.ethz.ch/quakeml/ , or http://earthquake.usgs.gov/earthquakes/feed/v1.0/quakeml.php)

3.10 3D models

WP3 (task 3.4) will deliver technical requirements for serving 3D geological models.

The way to develop 3D geological models depends on the complexity of geology and on the available data. In the context of basin environment it may be sufficient to interpolate horizons. Where geology is layered 2D methods are adapted to construct horizons and cross-sections. But geological bodies may have complex geometries, depending on their history, and then other methodologies are needed to specify the geological unit volumes and the faults.
According to the geological context, the 3D model is described by layers (2D surfaces defining volumes – often known as 2.5D) or by voxels (3D pixels) or by 3D surfaces defining volumes with complex shapes.
Often, once a geological model is established, other parameters as hydrogeological, geothermal, geotechnical parameters must be taken into account.

Several tools have been developed to create and manage 3D geological models (GOCAD, GSI3D, Geomodeller, GDM-Multilayer, SeisVision, …) but each of them have their own data structure.
The European project GeoMol (http://geomol.eu/home/index_html) will provide consistent 3-dimensional subsurface information based on coherent evaluation methods and commonly developed criteria and guidelines across several countries (Austria, France, Germany, Italy, Slovenia and Switzerland). The challenge is to provide a 3D model of a geological unit (the “Molasse”) in the Alps. For this 3D modelling, the goal is to:

- Describe models using open standards formats
- Store models in databases
- Access to models through standardised web services (for query, extract, update,...)

Some standards data models can help:

- GeoSciML (see previous section)
- GST: Geosciences in Space and Time (http://tu-freiberg.de/fakult3/gy/mageo/projekt_gst.html), connected to GOCAD
- RESQML (see previous section about geology)

The conclusions of the GeoMol project will be of high relevance for the future EGDI implementation.

3.11 Energy resources model

WP3 has not identified this domain as a priority, but it is however in the scope of the European geological surveys. A INSPIRE data specification exist for this thematic domain.

The Energy Resources model is structured as four separate application schemas which are created to represent the different approaches to model Energy Resources. The Energy Resources Vector, Energy Resources Coverage and the Energy Statistics application schemas depend on the Energy Resources Base application schema, which provides a base set of common Energy Resource classes including coded values for the classification of fossil fuel, renewable and waste Energy Resources.

More information about the data model for Energy resources:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_ER_v3.0.pdf

3.12 Comments for EGDI about data models

For the selected EGDI thematic areas, INSPIRE has already specified European data models that will be implemented by the Member States to deliver the related data. As the scope of EGDI is the European level (and not the national or local levels) it is recommended:

- to use these data specifications presented above when possible,
- to improve them in coherence with international standards when addressing extra needs raised by specific use cases,
- to collaborate with the working groups defining the international standards.

About 3D models, there is not yet a common way to create, describe, store and display 3D geological models. The conclusions of the GeoMol project, with the requirement of sharing 3D information across several countries will be highly relevant for EGDI.

4 Vocabularies for semantic interoperability

Many properties of a data model have a value to be selected from a controlled vocabulary. The vocabularies (or code-lists) are a key point for the semantic interoperability. At the European level it is necessary to establish common code-lists and for the various providers to define the “mapping” between their terms and the European ones.

4.1 Existing code-lists, and registries

INSPIRE Registries

The EC has setup registries related to INSPIRE Data Specifications. These registries contain code-lists specified in the regulation and must be used to deliver European datasets.

Each code-list must have an identifier (e.g. for Lithology: http://inspire.ec.europa.eu/codelist/LithologyValue)

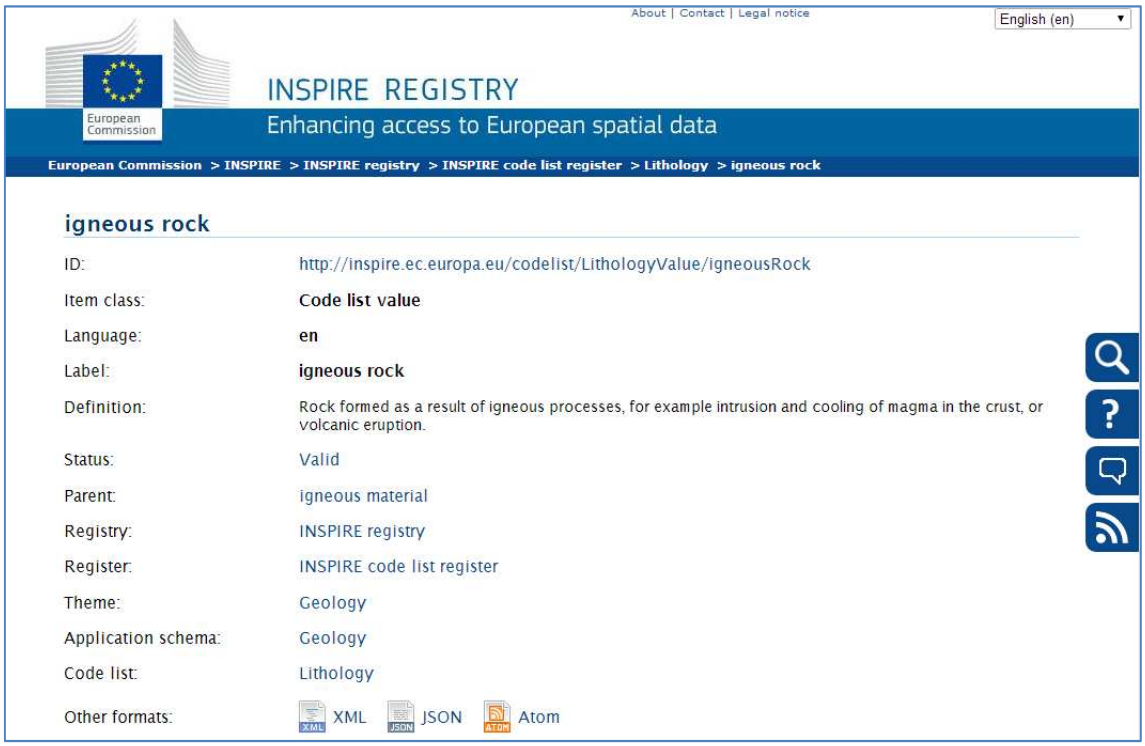
And each term of the code-list has:

- An identifier (an uri) (e.g. http://inspire.ec.europa.eu/codelist/LithologyValue/sedimentaryMaterial
- A label (e.g. sedimentary Material)
- A link to a “parent” (e.g. compound material)
- A list of narrower terms (if applicable)
- A definition
- A status (valid, invalid, retired, submitted, superseded)

The label and the definition could be provided in various languages

The content is delivered according to various formats (XML, JSON, Atom)

An example of lithology (value= igneous rock):



More information about the INSPIRE Registries:
http://inspire.ec.europa.eu/codelist/

EIONET – GEMET Thesaurus

EIONET (European Environment Information and Observation Network) has setup the GEMET thesaurus (General Multilingual Environmental Thesaurus)
Several GEMET themes are of interest for EGDI domains: climate, energy, disaster, natural dynamics, resources, waste, water

Each term/concept has an identifier (an uri), a label, a definition, a scope note, a link to broader terms and narrower terms, a link to one or more themes and relations with other existing vocabularies. Labels and definitions are available in many languages

The GEMET data are exposed through web services (see next section)

More information about GEMET:
http://www.eionet.europa.eu/gemet/

IUGS/CGI and OneGeology

The CGI (Commission for the Management and Application of Geoscience Information; http://www.cgi-iugs.org/), is a Commission of the IUGS (International Union of Geological Sciences). In this commission the task of the Geoscience Terminology Working Group is to

develop, review, adopt, publish, and steward vocabularies and associated documentation for use in geoscience information systems.
http://www.cgi-iugs.org/tech_collaboration/geoscience_terminology_working_group.html

Many vocabularies related to Geology and Mineral Resources are available. The Geology vocabularies are encoded using the SKOS format and are available for a user through this web site:

http://srvgeosciml.brgm.fr/eXist2010/brgm/client.html

The CGI vocabularies are also exposed through web services (see next section)
These vocabularies are used in the OneGeology project (http://portal.onegeology.org/)

4.2 Web services to access vocabularies

GEMET

GEMET's data is exposed through the Web for remote applications using XML (RDF/SKOS), HTTP and XML/RPC.

GEMET manages “concept” with the following properties (from SKOS: preferred label, definition, scope note, none preferred labels, example, acronym label.
Several functions (or methods) are available to use the GEMET vocabularies:

1 getTopmostConcepts: Get the top concepts of a thesaurus
2 getAllConceptRelatives: Get a list of all concept relatives.
3 getRelatedConcepts: Get a list of all concept with a given relation to the object.
4 getConcept: Get a concept by a known URI.
5 getAllTranslationsForConcept
6 getConceptsMatchingKeyword: Get a list of concepts matching a keyword for a particular thesaurus.
7 getAvailableLangs: Return the languages a concept's preferred label is available in.
8 getAvailableThesauri: Return all the thesauri uris, the service knows of

About GEMET web services:
http://www.eionet.europa.eu/gemet/webservices?langcode=en

IUGS/CGI and OneGeology

A web service is also available to access the CGI vocabularies (related to Geology) providing methods very similar to those developed for the GEMET Thesaurus.
Access to CGI web service:
http://srvgeosciml.brgm.fr/eXist2010/brgm/CGI201012uriindex.html

INSPIRE

Only a web client is currently available to access vocabularies. It is then recommended to check what INSPIRE will implement to complete this access.

4.3 Comments for EGDI about vocabularies

There are several main aspects related to vocabularies for EGDI:

- The governance of the content:
  The INSPIRE code-lists will probably be not complete enough to describe “real” use cases. For example for Lithology, the code-list has only 29 values, other values are described in the CGI Lithology code-list. So there will be a need to coordinate with CGI how the INSPIRE code-lists are updated.
  A Working group **must** be in charge of the definition of the content (the concepts and their relationships), of the validation of the various versions. It is also recommended to keep the link with other vocabularies in the same domain
- The technical point of view:
  - To manage the vocabularies:
    - To agree on an encoding (SKOS/RDF seems to be the standard)
    - To provide tools for the experts to define, comment vocabularies
  - To make the vocabularies available
    - for the users (similar to GEMET or CGI web clients)
    - for applications, portals through web services: the methods of such a web service must be specified (the GEMET Web service could be a good starting point or even a model, it is also recommended to check what INSPIRE will implement – only a web client is currently available)
- A clean mechanism to validate XML/GML (in WFS) reusing vocabularies for attributes values.

5 Services

The services identified in this section come from:

- The services requested by the INSPIRE Directive
- The services identified and implemented by previous and on-going European projects

The services more related specifically to infrastructure will be addressed in the deliverable D4.3 “Report on infrastructure needs”. In this category are included access control services (authentication, authorisation …), security, monitoring the Quality of Service, services for replication, for the use of the cloud technology, caching, tiling, …

EGDI to be compliant with the INSPIRE infrastructure must implement at least 3 types of web services (named **Network Services**):

- Discovery services to provide the users with the possibility to search for datasets and web services through their metadata describing them
- View services to display the data on a portal or an application
- Download services for the users who have discovered and viewed the data to get them and process them with dedicated tools.

INSPIRE has also specified rules to implement **Spatial Data Services** (services not yet specified but will process INSPIRE data). If providers want to connect this type of services (could be also named “thematic services”) to the INSPIRE infrastructure, they have to follow some technical rules.

5.1 INSPIRE Services

With the Network Services, INSPIRE addresses the general “Publish-Find-Bind” requirement.

1. **Discovery service**

INSPIRE Discovery Services allow users and computer programs to search for spatial datasets and services based on their metadata records

The discovery service specifications are based on the OGC™ Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile for CSW 2.0. It defines the following operations:

- Get Discovery Service Metadata: Provides all necessary information about the Discovery Service and describes service capabilities;
- Discover Metadata: Allows requesting INSPIRE metadata elements of spatial data sets and services from a Discovery Service;

- Publish Metadata: Allows editing of INSPIRE metadata elements of resources in the Discovery Service (push or pull metadata mechanisms). Editing meaning insert, update and delete;
- Link Discovery Service: Allows the declaration of the availability of a Discovery Service for the discovery of resources through the Member State Discovery Service while maintaining the resource metadata at the owner's location.

Search capabilities

The following list is the list of the main search criteria that a discovery service must implement in the discover metadata operation (= to the getRecord operation of the CS-W specification):

- keywords
- topic category
- resource title, resource abstract, resource type and resource identifier
- geographic bounding box
- temporal extent
- scale or spatial resolution
- responsible party
- language
- creation and publication dates
- service type

Discovery client approach

There are two possibilities for a client to be connected to a group of national catalogues:

- The discovery client knows all the catalogues, and is able to send a request to each of them:
  - Advantages:
    - Searches can be processed by the client: so the client can decide by its own how the search is operated.
    - The response time of a single search request may be more predictable as no hidden requests to third party catalogues are involved.
  - Disadvantages:
    - Every client has to determine the Discovery Service topology from time to time.
    - The searches must be processed by every client (it is not transparent to the client).
    - Discovery Services which are not directly accessible (e.g. running behind a firewall in an intranet) cannot be accessed.

- The discovery client is connected to only one "European Catalogue" which is connected to all national catalogues
  - Advantages:
    - The Discovery Service must only know its direct "child-catalogues".
    - Discovery Services behind a firewall can be accessed.
    - Searches don't have to be processed by every client.

- Disadvantages:
  - More enhanced query request and response structures are needed.
  - Every Discovery Service that provides access to federated catalogues must process searches.
  - The response time for a single request may be less predictable as possibly hidden requests to (potentially slow) third party catalogues are involved and may infringe the QoS requirements. To speed-up very slow responding remote Discovery Services a Discovery Service may harvest their content from time to time (creating an entire local cache of the metadata) and perform searches locally by filtering on all cached results of such a catalogue.

More information about the INSPIRE discovery services:
http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_DiscoveryServices_v3.1.pdf

2. View service

INSPIRE View Services allow users and computer programs to view spatial datasets.

The view service specifications are based on the ISO Web Map Service Specification and also uses the OGC™ Styled Layer Descriptor Profile (OGC SLD), and the OGC™ Symbology Encoding Implementation Specification (OGC SEIS). It defines the following operations:

- Get View Services Metadata: Get metadata about a specific view service;
- Get Map: Returns a map for a specified area;
- Link View Service: Allows the linking of view services together.

An INSPIRE view service may also be implemented as a OGC™ Web Mapping Tiling Service (WMTS)
This service is very well known and implemented by many projects.

More information about the INSPIRE view services:
http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_ViewServices_v3.11.pdf

3. Download service

The download services enable copies of spatial datasets, or parts of such sets, to be downloaded and, where practicable, accessed directly.
INSPIRE specifies two types of download services:

- A service able to download pre-defined datasets contained in multiple physical files
- A service able to provide a direct access to the data, with the possibility to define queries to get a selected subset of the data

The mandatory operations to implement for both types are:

- Get Download Service Metadata: provides all necessary information about the service, the available Spatial Datasets, and describes the service capabilities
- Get Spatial Dataset: allows the retrieval of a Spatial Dataset
- Describe Spatial Dataset: returns the description of all the types of Spatial Objects contained in the Spatial Dataset.
- Link Download Service: allows the declaration, by a provider, of the availability of a Download Service for downloading Spatial Datasets or, where practicable, Spatial Objects,

And for the optional direct access service, two more operations:

- Get Spatial Object: allows the retrieval of Spatial Objects based upon a query
- Describe Spatial Object Type: returns the description of the specified Spatial Objects types

For these two operations, the search capabilities shall also be implemented. These capabilities include the ability to search by:

- URI of Spatial Dataset
- Key attributes of spatial objects, including URI and date/time of update
- Bounding Box
- Spatial data theme
- Combinations of the above

The download pre-defined datasets is implemented using the Atom syndication format.
The direct access download service is implemented using the ISO 19142 Web Feature Service and ISO 19143 Filter Encoding.

The download service can be implemented using a Web Coverage Service (OGC WCS) if the data are gridded, and the use of Sensor Observation Service (OGC SOS) is also an option.
For INSPIRE the download service should deliver data according to a GML application format, which is based on GeoSciML, EarthResourceML, and GroundwaterML, that means XML format. But for some use cases it is requested to deliver the data in other formats (as Excel spreadsheets or shapefiles): in this case a common structure must be specified (as they are “flat files” and not complex features). And the way to transform the complex features managed by the INSPIRE model into a flat structure will have to be specified according to the use cases.

More information about the INSPIRE download services:
http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_v3.1.pdf

4. Spatial Data services

Spatial data services are the operations which may be performed, by invoking a computer application, on the spatial data contained in spatial data sets or on the related metadata.

The INSPIRE directive imposes to create discovery metadata for all spatial data services operating upon INSPIRE data.

INSPIRE defines various types of Spatial Data Services according to the way to use them (and not according to their functions):

- discoverable spatial data services are services described by INSPIRE metadata
- invocable spatial data services are services being discoverable and services which can be invoked by another service or application (a resource locator is available, defining an access point and is an URL; the service is well documented)
- interoperable spatial data services are discoverable and invocable services that follow specific requirements, for instance on the applicable spatial data sets.
- harmonised spatial data services are discoverable, invocable, compliant with interoperability arrangements and do fulfil additional requirements (as Quality of Service).

It is recommended that EGDI services when they process data of INSPIRE themes are at least Discoverable Spatial Data Services (described by INSPIRE metadata.

More information about the INSPIRE Spatial Data Services:
http://inspire.jrc.ec.europa.eu/documents/Spatial_Data_Services/Draft_TG_for_INSPIRE_SDS_2.0.1.pdf

5.2 Services identified in European projects

This section describes the services identified into previous or on-going European projects related to geology. The INSPIRE network services (discovery, view, download) often implemented in these projects are not described again in this section.

From several projects, we can identify services to the users for:

- Providing diagrams – time series (EuroGeoSource, eWater, …)
- Computing statistics (OneGeology-Europe, EuroGeoSource) – possibility to use the OGC Table Joining Service that enables joining tabular and geographic data.
- Delivering gridded data through a Web Coverage Service (EarthServer) and using 3D models
- Combining datasets as gridded data (EuroGEOSS) to address multi-disciplinary requirements
- Web Processing Services (EuroGEOSS), using the OGC WPS specification to encapsulate processes in a standard interface
- Workflow management (EuroGEOSS)

5.3 Search possibilities

Query interface: OpenSearch

A key point for EGDI is to expose European datasets, services, and documents … to the European community. One possibility is to setup a metadata catalogue based on IN-SPIRE/ISO metadata with some “queryables” elements.

Another option is to setup a catalogue exposing OpenSearch interfaces. Then search clients can use OpenSearch description documents that indicate how the search client should make search requests. OGC has extended the OpenSearch standard with geospatial and time extensions and it is a possible profile for the CS-W 3.0 part 4. The results are provided in various formats (HTML, Atom, XML/RDF, KML, JSON,WKT, …) directly integrated in web browsers.
There are implementations of OpenSearch in several tools as Geonetwork, OpenLayers, Gi-CAT (base of the GEOSS Broker).

Search engines:

1. Solr:

Solr is a standalone Open Source enterprise search server with a REST-like API, able to index and retrieve any document in various formats (XML, JSON, CSV or binary). It uses an existing search library (from the Apache Lucene project) and extends it.
Its major features include powerful full-text search, hit highlighting, faceted search, near real-time indexing, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search.

2. Elasticsearch:

Elasticsearch is a flexible and powerful open source, distributed, real-time search and analytics engine. Architected from the ground up for use in distributed environments where reliability and scalability are must haves, Elasticsearch gives the ability to move easily beyond simple full-text search.
It is document oriented storing complex real world entities as structured JSON documents. All fields are indexed by default, and all the indices can be used in a single query, to return results.
Elasticsearch uses Lucene internally to build its state of the art distributed search and analytics capabilities.

5.4 Comments about services for EGDI

As EGDI will deliver data for Europe, it is recommended to implement the INSPIRE Network services (discovery, view and download) as described in the previous sections. It is also recommended to follow the Spatial Data Services rules for all services EGDI wants to publish on the INSPIRE infrastructure.

For an EGDI metadata catalogue harvesting national catalogues to identify only interesting datasets and services for the European level, it will be necessary to find a way to distinguish the relevant metadata in the national catalogues from the others.
A very simple way for the providers to do it is to add a specific keyword (as “EGDI”?) in the keywords list of the metadata records, so that the EGDI harvester will select on the metadata containing this keyword.

There are also other interesting interfaces or search engine to look at (as OpenSearch and Solr) as they could offer a large audience to European geological products.

More technical services, related to the implementation of the infrastructure will be described in the deliverable D4.3 Report on the infrastructure needs.

6 Portal specifications

6.1 EGDI Resources portal and thematic portals

To address the infrastructure needs, it seems necessary to think about:

- An “EGDI Resources” portal which gives access for the Geological Surveys members to all resources shared:
    - A catalogue of datasets and services metadata describing and giving access to all datasets and services that Geological Surveys will deliver for European needs
    - Registries containing various elements as:
        - List of portals of existing projects
        - Components (catalogues, software, …) from projects that could be re-used
        - List of standards and best practices recommended by EGDI for a project development and implementation
        - List of existing vocabularies and related web services
        - Definitions of objects managed in various projects (similar to the INSPIRE Feature Catalogue)
        - Common rules to portray features (and SLD files)

    - Brokers: set of tools to discover and access data from various projects and using various formats
- Several thematic portals as it is not possible to develop only one portal to address all the user requirements related to many domains covered by the surveys. Specific portals, or applications, have to be developed for specific groups of users.

The EGDI Resources portal should have a dedicated **search engine** to provide access to different types of resources for a specific domain: as an example a user could search for a Web Map Service delivering a geological map on a specific area. Once the URL of this service is delivered the user can use it in a project (in another portal or another catalogue). It should be also possible or example to get a SLD file (very time consuming to create it) to portray data according to a standard way (defined by EGDI)

The EGDI Resources portal should provide also:

- A metadata viewer: when a dataset or a service is discovered the portal should display the whole related metadata record
- A data viewer: it could be either with a preview picture or by using a Web Map service

The EGDI Resources portal should be connected to other EU portals.

6.2 Connection with the EU Open Data Portal

The goal of this portal is to give access to EU data; it is and will be a single point of access to data produced by institutions (for data free to use and re-use, link and redistribute for commercial and non-commercial purposes.
The DCAT-AP (Data Catalogue application profile for data portals in Europe) is an initiative jointly chartered by DG CONNECT, DIGIT, and the OP, in the framework of the EU ISA Programme.

6.3 Thematic portals

It seems difficult to design one portal for the various thematic domains covered by geosciences in Europe (Geology, groundwater, mineral resources, natural hazards, …). Each of them has its specific requirements, but there are some common components:

- A search function to look at datasets and services, but also at related documents
- A map viewer with well-known functions: zoom in-out, pan, change layer transparency, define the layer rank, get metadata (including licence information), get information about an object on the map, gazetteer, get the legend, to view an external WMS (providing its URL)…
- Specific functions according to the data theme:
    - to provide statistics or diagrams
    - to select a subset of data according to properties

    - to display the subset of selected data
    - to download the subset of selected data
- The multilingual user interface, and the possibility to get metadata titles, descriptions, keywords in the selected language when they are available

To develop these thematic portals, it is recommended to specify and implement some components and to re-use them (a search interface, a map viewer, a metadata viewer, …).

7 Conclusions

This report provides input for EGDI implementation about several components of the system:

- Metadata describing datasets and services with a proposal to use INSPIRE metadata (based on ISO 19115/19119). This option coupled with the discovery service implementing the OGC CS-W specification will offer a connection to EPOS (through the CERIF metadata). EGDI should also be able to deliver open data to the EU Open Data Portal by providing DCAT AP elements,
- Standard data models are available through the INSPIRE Data Specifications for most of thematic domains identified as a priority, probably to be extended (thanks to the international standards) to address more use cases,
- A standard data model for 3D geological models is still in development, and the GeoMol project should provide relevant possibilities,
- The vocabularies or code-lists are also a key point for semantic interoperability. INSPIRE and several organisations provide values but there is a need for governance of the content and an agreed specification for the web service able to provide access to these vocabularies for users, portals and applications,
- A few services are mandatory to make the infrastructure deliver some products at a minimum: discover, view and download data. But more thematic services are necessary for EGDI: searching data according to properties, computing statistics, providing diagrams (time series), combining data (often using gridded format),
- The more technical services will be described in the deliverable D4.3 “Report on infrastructure needs”
- About the portals: it is recommended to develop an EGDI Resources portal to provide access for experts users to all components produced by EGDI partners and projects; and to develop dedicated portals for specific groups of users.