EURARE IKMS: AN INTEGRATED KNOWLEDGE MANAGEMENT SYSTEM FOR RARE EARTH ELEMENT RESOURCES IN EUROPE

Daniel CASSARD[1], François TERTRE[1], Guillaume BERTRAND[1], Frands SCHJØTH[2], Jørgen TULSTRUP[2], Tjerk HEIJBOER[2], Jouni VUOLLO[3]

[1] Bureau de Recherches Géologiques et Minières (BRGM), Orléans, France
[2] Geological Survey of Denmark and Greenland (GEUS), Copenhagen, Denmark
[3] Geological Survey of Finland (GTK), Rovaniemi, Finland

Corresponding author: D. Cassard (d.cassard@brgm.fr

Abstract

The EU-FP7 project EURARE 'Development of a sustainable exploitation scheme for Europe's Rare Earth ore deposits' will develop an operational data management distributed system based on high-level interoperability standards. Using advances made in EU-FP7 former projects such as OneGeology-Europe, ProMine and EuroGeoSource, ongoing projects like Minerals4EU in terms of database structure, harvesting systems, web services, metadata management, and integration of non-structured information, and InGeoCloudS project in terms of cloud computing, EURARE will contribute to implement the standards of a European geoscientific data infrastructure defined in the EU-FP7 EGDI-Scope project. The objective is to develop an Integrated Knowledge Management System (IKMS) allowing to easily combine rare earth elements (REE) information related to primary and secondary mineral resources and to provide end-users with all the available information from primary sources to new processing techniques and to waste streams. The IKMS will thus be designed as a fully-fledged extension of the EU-MKDP (EU-Minerals Knowledge Data Platform) currently under development within Minerals4EU and which aims to become the future European Mineral Resources (s.l.) data infrastructure. This assures an effective and sustainable system designed for facilitating data updates and maintenance and allowing a full access to information related to the whole REE resources value chain.

1 – Introduction

Data related to raw materials, either metallic, industrial or construction materials, primary or from wastes (mining and industrial) are, most of the time, available in Europe. However, they are often scattered amongst a variety of institutions, including governmental agencies, universities, NGOs and industries. These data are often stored in databases with their own design/architecture and vocabulary, making any attempt of merging – even partially - in view of a compilation difficult and time consuming. The problems regarding availability, quality, organization, accessibility and sharing of data are common to a large number of policies and are experienced across the various levels of public authorities. Solving these problems requires measures that address exchange, sharing, access and use of interoperable spatial data and services both at national and European levels. This is the aim of the INSPIRE Directive[1], but its

implementation in the Member states is just starting and achieving those objectives remains a major challenge.

The development of the IKMS (Integrated Knowledge Management System) aims to give a simplified, user-friendly and efficient access to all available and new data related to REE (Rare Earth Elements) from national geological surveys, scientific institutes and universities, relevant industries and professional organizations, as well as from former European projects such as ProMine[2,3] (information on both mineral deposits and anthropogenic concentrations resulting from mining and downstream activities) and EuroGeoSource[4,5] (information on energy and mineral resources, extraction locations, production, reserves). The system is also designed to accommodate and manage semi- and non-structured data (e.g., syntheses and statistics in the form of graph charts, time-series related to reserves and resources, secondary resources and exploration, European REE market survey and raw material demand, REE resource exploitation technologies including ore beneficiation, extraction technologies, end product development and waste management practices…).

The role of the IKMS is to provide the end-user with a seamless access to the whole value chain from REE deposit exploration, mining and extraction of ore, beneficiation and extraction technologies to treatment of end-of-life products and the generation of "new" materials, with the ability to combine all spatial and non-spatial pertinent information in a single reference system.

2 – IKMS general architecture

The principles of the IKMS architecture follow the Implementing rules of the European INSPIRE Directive[1] to make data and services interoperable across Europe over a distributed infrastructure.

The open architecture (figure 1) defines various components connected together to build the IKMS, taking into account state-of-the-art developments to enhance the performance, stability, sustainability and user friendliness of the system. The connections between the components are specified by standardized interfaces.

The main components of the IKMS architecture are the followings:

- Data on primary and secondary resources from databases provided by geological surveys and former European projects such as ProMine and EuroGeoSource. These data are delivered through INSPIRE compliant web services and according to INSPIRE (plus extensions) data model[6] and ERML v.2 data model[7,8,9,10,11].
- A system which gives access to available pertinent layers (e.g. geological, geochemical, geophysical, geographical, land use maps…) provided as interoperable services that are registered in a metadata catalogue (accessible via an INSPIRE discovery service).
- New data and information delivered during the project by other work packages (statistics analyses, analyses of REE supply and demand in the EU, stocks, flows, including trends in what products are put on the market and their composition,

analyses of the composition of waste products and wastes arising from pre-processing and their geographic location in EU) in various formats will be accessible through the portal. This applies also to any 'open data' of interest for the project. This semi- or non-structured information will be enriched by the provider with a metadata record (e.g., Dublin Core type, see below), then processed and indexed, first extracting relevant information from the documents (e.g., named-entity recognition on atomic elements such as names of locations, expressions of time, thematic specific terms…), then classifying this information according to three facets (spatial, temporal and thematic) and creating indexes for each facet. The way to display the list of selected documents will take into account these three facets.

- Besides a map viewer, the IKMS also includes a search engine for all data, layers and documents delivered by the project and all external pertinent documents.

3 – IKMS detailed architecture

In order to increase the value of the work done in former EU-FP7 project, and also to minimize new developments, it was agreed to reuse the Central Database of the EuroGeoSource project in EURARE. To achieve this goal, there was a need to both update the structure of the database to be ERML 2.0 compliant, and add the support of wastes and chemical analyses. There was also a need to update the database to use the final code lists (or vocabularies) coming from INSPIRE MR specifications which were themselves improved and validated at an international level by the International Union of Geological Sciences (IUGS), Commission for the Management and Application of Geoscience Information (CGI) and the Geoscience Terminology Working Group (GTWG). Finally, the EuroGeoSource cookbook used to help the partners to serve their data has been updated accordingly to the new structure of the database.

3.1 - A professional architecture

The IKMS adopted a professional architecture with a harvesting system and a diffusion system (figure 2). The system is divided into three parts:

- The national level from which ERML 2.0 compliant WFS' (Web Feature Service) deliver data to the harvesting system either from (already) EuroGeoSource structured databases or directly from other data sources in each country;
- The central harvesting system which regularly reads data from the national WFS' and stores it in the Harvesting Database. Data is subsequently being delivered to the central diffusion system through a database synchronization mechanism. The Harvesting Database is optimized towards reading data from the national level and delivering this data to the diffusion system;
- The central diffusion system which is updated with data from the harvesting system at regular intervals and makes this data available to users through the IKMS web portal. The Diffusion Database has specific optimization to offer the better experience to the user for the delivery of the data and the computation of on-demand services.

The Information Factory is a software component able to process data (indexation, filtering, descriptive/exploratory statistics computation, automatic report generation…).

3.2 - The national level provider system

Each participating survey will have a National Provider Database which will have the same architecture as the Central Harvesting Database. This database is a PostGreSQL database, which build upon the schema that was used in the EuroGeoSource project, extended by the requirements from EURARE (including mining wastes and geochemical analyses). This Provider Database will contain the data of a survey that was mapped from that survey's own database using the INSPIRE vocabulary defined in the code lists and data types. This mapping can be done using an ETL (Extraction-Transformation-Loading) tool such as GeoKettle. The data from this database will then be transformed to an INSPIRE compliant web feature service (WFS) using the DeeGree3 framework and toolstack.

3.3 – The harvesting system

The harvesting procedure is based on an updated EuroGeoSource harvesting procedure. The harvesting from the National Provider/project WFS' is done with an ETL tool run in a scheduled job. Aggregation and other manipulation of data in the harvesting system, if needed, will be performed by procedures within the PostGreSQL-database.

The possibility to set up an incremental harvesting system, which would avoid a total refreshment of harvested data, is still under consideration. The problem here is not of a technical nature but depends on the fact that some data may not have a time stamp.

3.4 – The Harvesting Database and the INSPIRE MR data model

The Harvesting Database is built by extending the existing EuroGeoSource Central Database with new tables. Additions and changes to the INSPIRE MR data model include:

- the Mining Waste extension from INSPIRE, and the modification of the Earth Material feature used in EuroGeoSource in order to include other materials;
- parts of INSPIRE Draft Guidelines for the use of Observational Measurements and Guidelines for storing geochemical measurements of Earth Material specimens;
- the Mined Material part of the Mine extension from INSPIRE v.3;
- the Supergene Processes from the Occurrence extension from INSPIRE v.3.

The delivery of data from the harvesting system to the diffusion system is carried out by sending SQL update scripts.

3.5 - Vocabularies

Vocabularies used in EURARE are based on INSPIRE MR code lists. These code lists, since their initial publication in the Data Specification on Mineral Resources – Draft Guidelines[6] have been submitted to the IUGS/CGI/GTWG (cf. supra) for validation at the international level and have thus been reworked and improved. During this process, it was ensured that all the needs of the EURARE project regarding REE were taken into account. This led to the improvement – before their final acceptance - of several code lists such as the Commodity Code Value, the Deposit Group and the Deposit Type code lists.

3.6 – The Diffusion Database

The Diffusion Database will be a copy of the Harvesting Database, which will be specialized for data delivery and services that carry out computations based on the stored data. This database will be kept updated with the Harvesting Database using database synchronization. This synchronization will rely on SQL scripts forwarded from the Harvesting Database to the Diffusion Database. The specialization of this database will aim to optimize the response time of the services proposed to the users.

The services on top of the Diffusion Database are:

- Simple map visualizations using WMS (Web Map Services).
- EURARE's dedicated services: these services are described in a dedicated paragraph below.
- Data download: in some case (and depending on the access rights to the data) the data will be available for download so that they can be processed in other applications (e.g., desktop GIS…).
- Search facilities: these facilities will allow the user to search in the whole database and documents.

3.7 – The non-structured data

Beside the Diffusion Database, an additional interface will allow experts to feed the knowledge base with non-structured data. These non-structured data can be of different types (reports, synthesis notes, thesis…). This interface will allow the expert to add a document (the document will be part of the IKMS and retrievable within the IKMS) or a link to an existing document in some place accessible via the Internet (the document will stay on its original server; the end-user will have to follow a link to retrieve it). To be able to integrate it into the knowledge base, the expert will have to include metadata to this document. These metadata will be based on Dublin Core (http://dublincore.org/) with a specific extension for EURARE (at least quality of the data). Dublin Core Metadata Element Set has 15 elements covering the most important properties to describe a document (title, creator, subject, coverage – temporal or geographical…). These metadata will then be used in the search facilities to retrieve the documents.

3.8 – The search facilities

The search facilities will be based on a search engine indexing the Central Diffusion Database, the documents corpus (non-structured data) and some external databases (if any). A user interface will allow the end-users to retrieve data from the whole IKMS using a simple input (Google-like search, a simple sentence will search in the whole IKMS), or using some specialized interfaces (search specific concepts using their main attributes). Almost all the indexed concepts will have a geographic and temporal extents (coming from INSPIRE MR/ERML for the Diffusion Database, coming from Dublin Core metadata for the documents). These extents will allow the user to retrieve the most accurate response for his search.

3.9 – EURARE's dedicated services

In order to answer the end-users most common requests, some specific calculation services will be set up for EURARE on top of the Central Diffusion Database. As it is difficult to include LREE and HREE in the Commodity Code Value code list, a service can be developed which will classify each deposit as a HREE or a LREE deposit, taking into account the main commodity, the second one, etc. This would, for instance, allow plotting deposits on a map with distinct symbols (HREE or LREE), which could be useful for end-users as both groups of REE have different industrial applications. It would also be possible to do this from the geological or metallurgical point of view, as both classifications are not similar. Similarly, a service could also calculate the ratio LREE/HREE, provided that available data is accurate and numerous enough. A service can also be set up for plotting simultaneously known REE deposits (having REE among the main commodities) and deposits containing REE (REE present, but (totally) subordinate). For the latter ones, the selection can be made based on the presence of one or several REE containing mineral(s) in the paragenesis and/or in the mineralogy. This could possibly help to better define provinces or districts.

Finally, both harvesting and diffusion systems will be self-contained systems that can easily be moved to other installations/platforms. More particularly, the project will investigate how platforms like InGeoCloudS could be exploited for the system. This work will be realized in close collaboration with the EGDI-Scope and InGeoCloudS projects.

4 – Conclusion

The present work does not only represent an advance in terms of implementation of a distributed architecture and management of structured, semi- and non-structured data related to REE. It also contributes to the implementation of a European Mineral Resources Data Infrastructure and to bring concrete answers to several recommendations made by the European Commission. The IKMS contributes to the improvement of the standardization of mineral resources data across participating organizations, as suggested in the WP3 of the EC Communication “Making raw materials available for Europe’s future well-being proposal for a European innovation partnership on raw materials” (COM(2012) 82 final) and to the improvement of their availability. It also contributes to implementing land-use planning best practices as

suggested in the relevant report by Enterprise and Industry Directorate General (http://ec.europa.eu/enterprise/policies/raw-materials/files/best-practices/sust-full-report_en.pdf). The IKMS outputs will be appropriate for use at a variety of scales, from local/regional land use planning up to the pan-European level. It allows better communication with the general public and specific stakeholders within the mineral sector and facilitates more balanced and well-informed debates and decisions related to the raw materials (and notably critical ones) non-energy extractive industry. Finally, it contributes to the enhancement of minerals role and related industrial technologies for competitive growth as part of the Europe 2020 Strategy.

Figures and captions

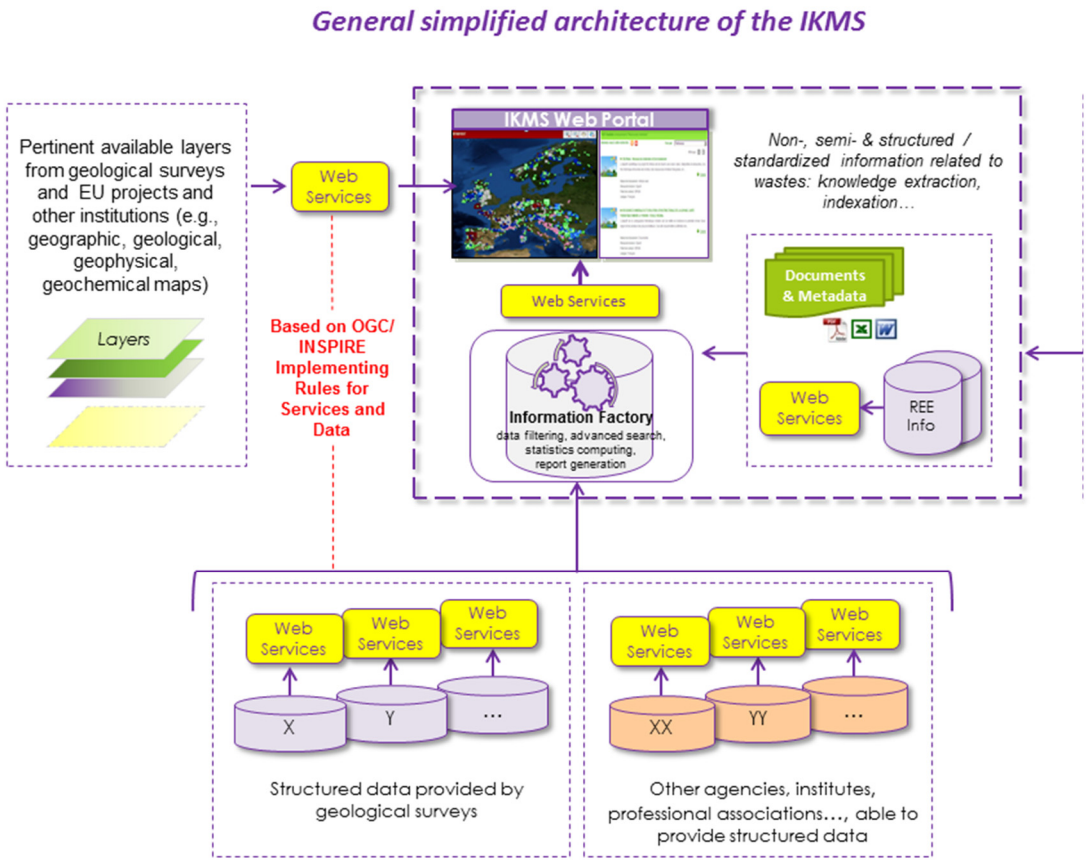

Figure 1: Simplified architecture of the EURARE IKMS (Integrated Knowledge Management System). The system is designed to accommodate both structured data from existing databases (data produced by geological surveys or external providers) and semi- or non-structured information produced by other work packages.

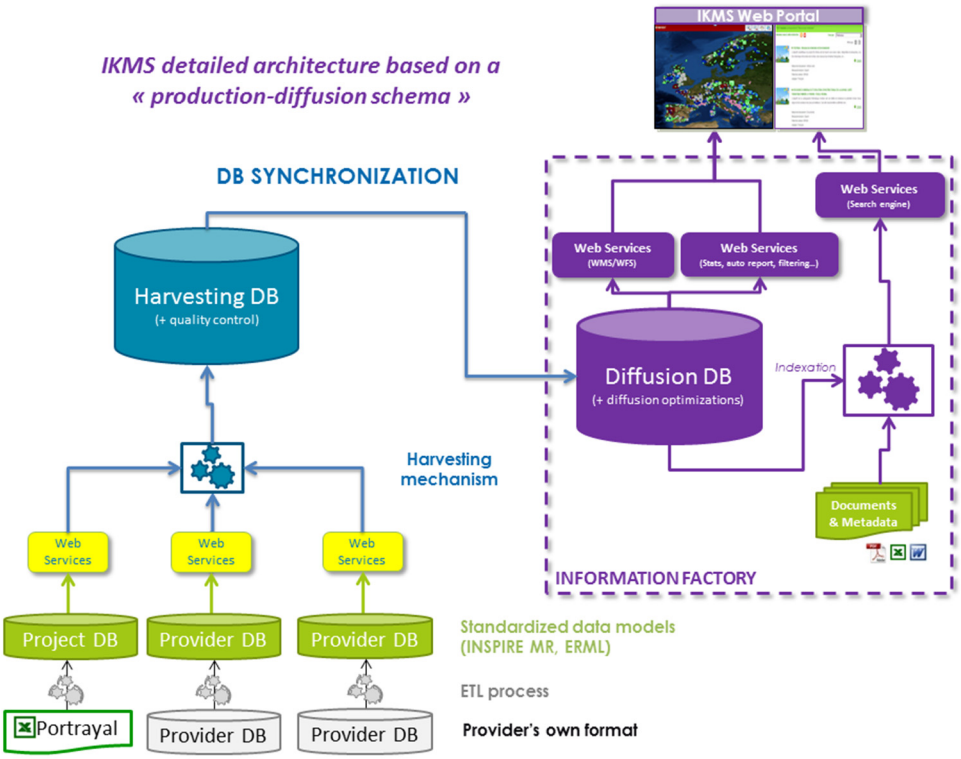

Figure 2: The IKMS detailed architecture showing the Harvesting Database, the synchronization process with the Diffusion Database, the Information factory and the indexation process of both structured and semi- and non-structured data. Note the possible use of Excel-type portrayals for feeding project databases.

References

1. *INSPIRE Directive (2007). http://inspire.jrc.ec.europa.eu/ ; http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF.*
2. *Serrano J.-J., Kauniskangas E., Cassard D. (2010). ProMine: architecture, portal and web services to provide a European mineral resources information system, 4th INSPIRE Conference. Krakow, 22-25 June.*
3. *Cassard D., Bertrand G., Billa M., Serrano J.-J., Tourlière B., Angel, J.-M., Gaál G †. (submitted). ProMine mineral databases: new tools to assess primary and secondary mineral resources in Europe. ProMine Special Volume. Springer Editor.*
4. *Gruijters S. and Fruijtier S. (2011). Using cloud computing to increase the usability of distributed Web GIS Systems. eChallenges e-2011 Conference Proceedings, 26-28 October 2011, Florence, Italy. Paul Cunningham & Miriam Cunningham (Eds.), IIMC International Information Management Corporation, 2011, ISBN: 978-1-905824-27-4, 8 p.*
5. *EuroGeoSource (2013). EuroGeoSource homepage http://eurogeosource.eu/. In the menu EuroGeoSource Deliverables: 1) Publications and presentations: http://eurogeosource.eu/deliverables/publications-and-presentations; 2) Public Reports: http://eurogeosource.eu/deliverables/public-reports*
6. *INSPIRE Thematic Working Group Mineral Resources (2013). D2.8.III.21 Data Specification on Mineral Resources – Technical Guidelines. 156 p. European Commission Joint research Center Publisher.*

http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_MR_v3.0.pdf

7. *EarthResourceML (2013). http://www.earthresourceml.org/earthresourceml/2.0/doc/ERML_HTML_Documentation/*
8. *Vuollo J., Cassard D., Simons B., Seymon A. (2010). The Earth resource data exchange model (EarthResourceML) – a tool for delivering INSPIRE and ProMine mineral resource data, 4th INSPIRE Conference. Krakow, 22-25 June.*
9. *Vuollo J., Simons B., Laxton J., Cassard D., Seymon A. (2012). EarthResourceML v.2.0 – an upgrade of the CGI-IUGS Earth Resource Data Model due to INSPIRE Data Specification. 34th International Geological Congress, Brisbane, August 2012, Abstracts on CD-ROM.*
10. *Cassard D., Serrano J.-J., Vuollo J. (2012). The use of the standard exchange EarthResourceML in the ProMine project. 34th International Geological Congress, Brisbane, August 2012, Abstracts on CD-ROM.*
11. *Raymond O., Vuollo J., Richard S. and CGI Interoperability Working Group (2012). Global Geoscience Data Transfer Standards - GeoSciML and EarthResourceML. eResearch Australasia Conference, Brisbane, Australia. http://conference.eresearch.edu.au/eres2012/posters-2012/*