# eEarth: Bridging the divided national geo-databases via multilingual web application.

*A. TCHISTIAKOV[1], J. Jellema[1], H. Preuss[2], T. Hernandez Diaz[2], B. Cannell[3], J. Passmore[3], T. Mardal[4], D. Capova[5], J. Belickas[6], V. Rapsevicius[6].*

1 - TNO-NITG, P.O.Box 80015, 3508 TA, Utrecht, NL, a.tchistiakov@nitg.tno.nl;
2-Bundesanstalt für Geowissenschaften und Rohstoffe (GE), 3 - British Geological Survey, UK;
4 - Polish Geological Institute, PL; 5 - Geofond, CZ; 6 - Lithuanian Geological Survey, LT.

## ABSTRACT

Access to geo-data repositories in Europe is effectively limited by national boundaries. The EC-funded project electronic Access to the Earth Through Boreholes (eEarth) is developing advanced software tools for multilingual access to the European national geo-databases. The new web services will include: a central web portal to the national geo-database applications, multilingual user interfaces to the national databases, "on-the-fly" translation facilities for standardised geo-scientific terms, GIS functions, access to geo-data via mobile devices, on-line data ordering and payment (http://eearth.nitg.tno.nl/).

## 1 INTRODUCTION

The national geo-scientific databases are the key elements of the national policies on the sustainable and environmental use of the subsurface. Geo-information is of particular importance for urban planning, where the engineering geological conditions are complicated by abundant subsurface infrastructure. European governments delegate the task of maintaining geo-scientific repositories to their national or regional geological surveys. The National Data Repositories (NDR) receive geo-data from a variety of sources: oil and gas operators, drilling firms, engineering and environmental companies etc and then are responsible for disseminating this data to national users. Publicly accessible databanks allow consultants and contractors to save on the considerable costs that can be incurred for storing and maintaining data and provides them with a rich source of information, an important consideration when assessing a new area.

International data exchange, however, is still poorly developed largely due to the following obstacles:

- The majority of geo-data web services, often financed by national governments, are aimed at national customers and therefore access to these services from abroad has not been a prime consideration in most cases. However, the intensive political integration within Europe together with globalization of the geo-data market raises questions regarding this "national" concept.
- Many web sites and database applications, allowing access to the borehole data catalogues, are available only in the host nation's language. Occasionally an English version is available in addition to the local language version. Thus many foreign clients find it difficult to gather even basic information about borehole data availability.
- The national legislations for geo-data dissemination do not regulate the data transition across borders, even within the EU.
- Different countries have been developing their own terminology for lithology and stratigraphy thereby making cross-border geo-data comparison and interpretation difficult.

In order to improve cross-border geo-data exchange, a consortium comprising eight European geological surveys, institutes and commercial companies have initiated the eEarth project. The project includes the following organizations:

- The Netherlands Institute of Applied Geoscience (TNO-NITG, NL)
- British Geological Survey (BGS, UK)
- German Geological Survey (Bundesanstalt für Geowissenschaften und Rohstoffe, BGR)
- Lithuanian Geological Survey (LTG, LT)
- Polish Geological Institute (PGI, PL)
- Geofond (CZ)
- Geodan Mobile Solutions (NL)
- Golder Associates (IT)
- The project objectives and structure.

The principal objectives of the eEarth project are:

- To increase availability, use and distribution of the European digital subsurface data by providing cross-boundary access to the digital geo-data collections in different EU languages
- To develop the international multilingual commercial services, including WEB GIS facilities, based on public geo-data stored in the national geo-scientific databases.
- To develop recommendations for new European standards for well data description, which can be used for other applications in future.

The project has three phases:

- Inventory / definition phase (month 1-9)
- Implementation and testing (month 10-15)
- Dissemination and Demonstration (month 16-18)

Phase 1 includes four main tasks:

- inventory of the national and international legislations for supplying state-owned geo-information through a public service.
- inventory of the software applications currently used for storage of geo-data from boreholes in the European countries.
- inventory of national geo-data formats and standards in order to establish a minimum common format for geo-data exchange.
- Design of the web-based multilingual services.

Phase 2 comprises three tasks:

- Implementation of the new multilingual data-access services at three national databases
- Demonstration of a mobile application for multilingual accessing a national database
- Demonstration of open source software solutions for developing web-based geo-database applications

Phase 3 principally concerns marketing of the portal and dissemination of project achievements.

## 2 THE PROJECT RESULTS

Results obtained during the first phase of the project, are presented below.

### 2.1 National and EU regulations on geo-data dissemination

In order to ensure that the eEarth cross border services comply with the legislation of the participating countries, an extensive study of the national and EU legal regulations on geo-data dissemination was completed. The research shows that no legal obstacles to cross-border dissemination of geo-scientific data exists in the participating countries, although sometimes the formal procedure for obtaining access to data is quite complicated.

The national regulations concerning geo-data dissemination have many common features. This creates a good basis for harmonisation of national legislation in the future. Notable differences centre on the granting of special access to groups of privileged users

Analysis of the EU legal initiatives confirms that the technical concept of the eEarth system is in line with overall EC policy regarding dissemination of environmental data and provision of access to public data holdings, using electronic means of data distribution.

### 2.2 Inventory of best practices in national borehole data management

The main task of this part of the project was to summarise the experience and approaches currently existing in the European Geological Surveys regarding description, collection, storage, retrieval, evaluation and distribution of borehole geo-data. This inventory allows identification of the most efficient and modern methods of storing and dissemination of borehole data.

In total we analysed nine national geo-databases of six European countries and the databases of eleven German federal states (Bundesländer). They typically contain 65-80% from the total amount of boreholes drilled in the countries. The sources of information are mainly the archives of the geological surveys and data obtained from drilling companies.

The borehole databases differ significantly in terms of hardware, operating systems, database software and data structure. The majority of the partners use UNIX (Solaris) and Windows operating systems in combination with an ORACLE database management system.

Coding standards are very diverse. Digital input of data is mainly handled by digital forms developed in-house, whereas commercial input systems, like ORACLE or Access forms, are rarely used.

Most database applications have GIS-functions enabling display of location maps. Shape files are normally used as the sources of geographic information. Profile construction, cross-sections and 3D-modelling functions are still under development by most of the eEarth partners.

The procedures for accessing geo-data are technically rather different. The majority of the national borehole databases (5 of 9) can be accessed externally. The web services normally include the display of location maps and metadata overview lists.

Electronic on-line data ordering and supply procedures are well developed in only a few countries, in others – data delivery by post is still the main option. Service costs and payment methods vary significantly, e.g. both prepaid charging methods and payment pro rata are used by the eEarth community.

From a technical point of view all the national borehole data management systems are quite advanced, however the partners continuously work on improvement of their display, evaluation and dissemination functions.

## 2.3 Development of the XML-standard for cross-border exchange of borehole data

Because geological practice has always been organised on a national basis, up till now there has been little incentive for the international exchange and harmonisation of geo-data. In order to design a common exchange format for this purpose, the data structures of the national geo-databases of six participating countries (United Kingdom, the Netherlands, Germany, the Czech Republic, Poland and Lithuania) have been analysed.

The new geo-data exchange format has been written in XML (eXtensible Mark-up Language). This language is both useful and popular in all kinds of Internet applications, where exchange of data takes place. Current practices in the oil- and gas community (WITSML) as well as in the mining world (XMML) have shown that XML is a powerful tool in connecting different data collections across disparate platforms. However implementation of XML for data exchange is only successful when all users agree on every detail of the format.

The selection of fields has been made against the background of four national exchange formats that have proven to be valid in geo-engineering and geo-environmental applications. As a result forty fields have so far been included, based on the following criteria:

- presence of these fields in the majority of the national database models;
- for metadata - the usefulness of a field as a search parameter for data set selection;
- for borehole interval data - frequency of use by users (relevance to national geo-markets).

## 2.4 The eEarth system technical design

The eEarth services will allow translations of the standardised geo-scientific terms on the fly, as necessary for metadata and for interval coded data. The demonstration version will include the English, German, Dutch and Polish, Czech and Lithuanian languages. The service will translate:

- The user interface texts.
- The codes and keywords of the borehole metadata.
- The standardized codes of the interval lithology

The architecture of the eEarth system is shown in Figure 1. The conceptual design comprises the requirements collected by the partners from their national and international users. The consortium concluded that a distributed system with individual components of the same look and feel would be the best option. It fits to the requirements and will minimize risk and effort. Below the main components of the system are described.

### 2.4.1 The eEarth Web portal

A central web portal will be the start point to eEarth services. The portal will be multilingual, encompassing the partners' European languages. This could easily be extended to incorporate additional languages in the future.

The eEarth web portal is proposed to be a relatively simple start page with only two necessary functions:

a) select a language (for further communication and the translation of borehole data)
b) select a country (which borehole data is needed)

The start page would provide links to a national survey's own database applications, where visitors could browse for the information required. From a technical point of view this means that the eEarth multilingual services will become a part of the national systems. The distributed system will:

- allow adjustment of the implementation of the software environment at the partner's organisation
- negate the risk should any element of the central portal be unavailable for whatever reason;
- eliminate the need to standardize coordinate systems and GIS software across the participating organizations, thereby reducing training, data conversion and maintenance overheads in subsequent years;
- allow implementation of the individual national payment schemes currently applied in the countries;
- minimise the interdependency of the partners, including the Web application support, after the end of the project.

### 2.4.2 Search and delivery of borehole information

When a user accesses a national eEarth web page he will be able to search for data in the national database via either a web-form or a web-based GIS interface. In both cases the data search is based on criteria defined by the user.

To enable the integration of GIS functions two scenarios for the web map server implementation are foreseen: 1) based on open source applications (i.e. UMN MapServer), and 2) based on a commercial software (i.e. ArcIMS by ESRI). Under both implementation scenarios, the GIS page will include a standard set of GIS functions, allowing geographic selection of boreholes from the national databases. A user will be able to select a single borehole via info-click or a number of boreholes by defining a selection area (rectangle) on the map. Subsequently borehole metadata can be displayed for the selected set.

Based on the metadata list, a user will be able to order descriptions of borehole layers along the borehole profile (interval data), when available. In some of the national database applications the interval data can be immediately displayed, in others the data is delivered either by post or e-mail.

Since the interval data is normally not delivered free of charge, access to them is restricted to registered users only. Therefore several special system modules will handle data ordering, user registration, and delivery (Fig. 1).

Electronic commerce functionality, allowing payment with a credit card, is only available at one partner organisation so far At the other national web database applications a user will use traditional methods of payment, such as money transfer via bank, etc.

### 2.4.3 Translation service

The translation service will provide on-the-fly translation of the user interfaces, borehole metadata and interval data (lithology, etc). The borehole data translation service is a quite complicated tool, which needs to translate coded geological data according to standardized geo-science dictionaries or thesauri. The main challenge results from the fact the translated meanings of many terms only partly overlap with their original definitions. In other words, "similar" terms can have quite different interpretations in different languages.

In order to provide multilingual services, a 6-lingual thesaurus stored in a database and software tools for data translation have been developed. The concept of a "distributed thesaurus", accepted by the consortium, presumes that one of the partners (i.e. Geofond) will maintain the central 'master' multilingual thesaurus (MMT) stored in its database. At the same time each partner geological organisation (PGO) will have individual (distributed) multilingual thesauruses (DMT) stored in translation tables within their national databases. Input and editing of individual terms in the MMT will be made by the partners by means of a special web application. Update of the translation tables in the national databases will be done via an export XML file that is generated by the MMT application.

The advantages of this concept are: 1) each PGO can implement the eEarth services using its own technology; 2) if the structure of the MMT changes, this does not directly affect the distributed thesauruses (DMTs) at the other PGOs.

### 2.4.4 Mobile service

The project will provide an engineer, operating in the field, with access to a national geo-database by means of a mobile handset (PDA). A "mobile" user will be able to visualize his location and order geo-data from the boreholes located in the area interest. The borehole will be selected by means of a GIS application developed particularly for a mobile handset.

The mobile services require special additions to the system design:

a) to serve mobile systems (special output and controls of the eEarth web pages) and

b) to accept locations of mobile systems (as an input to the borehole selection function).

Both additions are included into the design of the system, where (b) is supported by a position server at the national level only. It is assumed that mobile devices are able to send coordinates of their location in exact form so that the country selection of the start page can be omitted.

## 3 CONCLUSIONS

The concept of the eEarth system is in line with overall EC policy regarding dissemination of environmental data and the provision of access to public data holdings, using electronic means of data distribution. No legal obstacles to cross-border dissemination of geoscience data have been found on the national levels so far.

The new XML standard, developed by the project for exchange of borehole meta and interval data, will contribute to further harmonisation of geo-data in Europe. Data standardisation is a precondition for combining the national geo-data in a single pan European repository, which may be considered as an option for further unification of the geo-information in EU.

The national geo-databases contain 65-80% of the borehole data in the participating countries. Multilingual access to the national repositories via the Internet will significantly increase their added value, particularly for cross border projects. Access to the geo-information through mobile equipment will stimulate the use of the data by geotechnical and environmental specialists operating in the field.

Although many partners offer comparable services, the national borehole databases differ significantly in terms of data structure, database software, hardware, and operating system. Considering the implementation and maintenance issues the eEarth system has opted for a distributed structure. The eEarth conceptual design includes three main components:

1. A central multilingual web portal containing language selection and links to the national database applications
2. Master Multilingual Thesaurus (MMT) of geological terms maintained at one organisation and used by all the others for updating the dictionaries in their national geo-databases.

## 4 DISTRIBUTED NATIONAL MULTILINGUAL WEB APPLICATIONS, INDEPENDENT FROM EACH OTHER BUT HAVING SIMILAR INTERFACES AND FUNCTIONS.

This design allows a new partner easily join the multilingual services via the eEarth Web and access the MMT in order to include a new language. The national surveys that join the portal at a later stage will be required to purchase the associated training and support, thereby generating additional income for the project.

In order to insure high quality of this new type of the European geo-data service a comprehensive evaluation of functionality should be undertaken at all stages during development. This would include not just the member surveys, but also commercial partners involved with the project.

## 5 ACKNOWLEDGMENTS

We would like to acknowledge the EC eContent program for funding the eEarth project. Also we acknowledge the support given by the European geological surveys and institutes in sharing their knowledge and IT experience with the project.

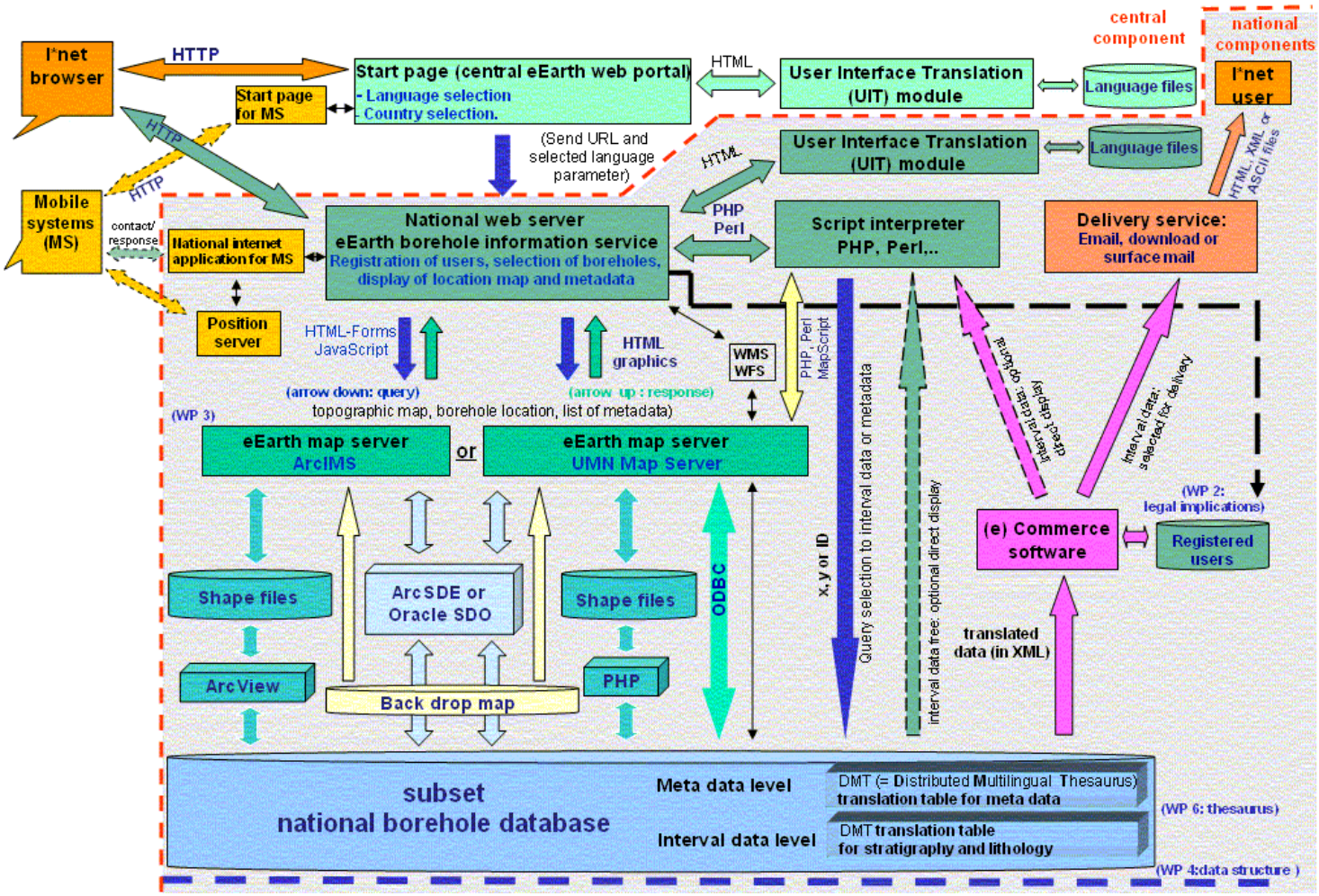


Figure 1: General diagram of the eEarth service implementation.