Provision of interoperable datasets to open GI to EU communities
An European example of SIM

Dr. Szabolcs MIHÁLY and Pál LÉVAI, Hungary

Key words: geographical information, spatial information management, European Union.

SUMMARY

A project named as entitled above and identified by GIS4EU is close to finish. The project is founded by the eContentPlus 2006 Programme of the European Commission, and includes 23 organisations of 10 countries in Europe. The project coordinator is Dr. Stefania de Zorzi of CORILA Consortium for Coordination of Research in the Venice Lagoon Systems in Italy. The Institute of Geodesy, Cartography and Remote Sensing is one of the partners involved.
The aim of GIS4EU project is to organise an infrastructure to share cartographic data and feature layers, in order to make them more accessible and to share the information more easily. This is to be accomplished without building a central database, rather by sharing information and data through standard services.
The project develops common data model and support tools for the assessment of common standards based on the 2007/2/EC - INSPIRE - Directive of European Communities though the following actions:

- Creating systematic approach and network covering all those involved in using geospatial data, i.e. data collector scientists, technicians, general users and other stakeholders;
- Overcoming critical issues and challenges in developing a common data model linking together all subjects involved in the decision making process;
- Generating a website for free testing and sharing of information and methodology;
- Improving communication and integration among the main actors and stakeholders at different levels of data productions and use;
- Generating new opportunities for sharing information and develop common technical approaches compliant with the INSPIRE Directive.

These actions are to create an operational validation of INSPIRE Implementing Rules and the GIS4EU project is to produce proactive suggestions addressing critical issues concerning operational implementation of the Directive.
The paper presents results of the project as an example of Spatial Information Management.

Provision of interoperable datasets to open GI to EU communities
An European example of SIM

Dr. Szabolcs MIHÁLY and Pál LÉVAI, Hungary

1. INTRODUCTION

While the EU is made up of 27 autonomous countries, many of its infrastructure systems – strategies, government policies, transport and communication networks – are interconnected. Base cartography datasets act as the fundamental basis of policy-making in areas such as national security, markets and environmental protection. It is therefore essential that such information is consistent, reliable and that it applies across all areas of the EU.

Currently data are collected and maintained mainly by individual public institutions, each of which has applied its own priorities and standards. Also, data are collected and stored at different levels of the respective organisations. This differing practice makes data integration on the pan-European level complex and difficult. The specific issues are:

- Fragmentation and diffusion of reference spatial data across Europe,
- Different formats of reference data – (i.e.: coordinate system, grid system, map projection, files and data base formats),
- Different classification of objects (e.g.: roads, cities, rivers, administration borders),
- Different cartographic presentations of the same objects,
- Lack of common rules for edge matching between data bases in different countries and regions,
- Lack of one common data model format,
- Different formats of data bases, different objects classification prevent data aggregation,
- Lack of access points for information about data sources (metadata), or to compile datasets from many sources and aggregate them to receive consistent data describing the whole European Community area,
- Lack of data access points available in national languages to allow above mentioned services,
- National laws applying different regulations for data access and disclosure including pricing policies.

These barriers can impede development and introduce obstacles in such areas as: spatial and environmental research, public and commercial innovative services, European policy and strategy creation. Factors such as national security and legal restrictions can also stand in the way of progress.

The European Union has made special efforts to address these concerns by approved the INSPIRE (2007/2/EC) directive (The Infrastructure for Spatial Information in the Community). The INSPIRE directive will establish a general framework for all Member States for the establishment of European spatial data infrastructure.

Under the INSPIRE initiative specific implementation rules are under development to introduce standards and requirements for national GIS data bases and SDIs which will be

binding on Member States. The most important factors from the perspective of reference data harmonisation will be metadata and data specification implementation rules.
Final metadata Implementation Rules (IRs) was established in 2008, final data specification was established in 2009 (Annex I) and are expected for 2012 (Annex II). During the Implementation Rules drafting phase pilots and prototype projects will play an important role, allowing operational validation of IR drafts. To achieve well defined, realistic, consistent and constructive Implementation Rules, operational validation is crucial.
Additionally, such projects will help data providers to prepare their spatial data infrastructures and data management, in order to fulfil INSPIRE requirements.

2. DESCRIPTION OF THE PROJECT

The project's main objective is to make spatial information more accessible, useful and exploitable by all users by providing consistent and aggregated base reference data.
The following specific objectives have been identified:

- To share base cartography in order for it to be accessible (INSPIRE compliance) and interoperable (cross-border/scale/language) for aggregation purposes,
- To provide reference data from the following themes: administrative units, transport network, hydrography, (Annex I of the INSPIRE Directive) and elevation – (Annex II of the INSPIRE Directive) in a common format to allow its aggregation.,
- To share data, without building a central database, but through service standards,
- To create the GIS4EU server which will enable access through standard services to data from different sources, by applying harmonisation,
- To analyse lack of information and dataset level,
- To operationally validate the INSPIRE Implementation Rules and make suggestions in order to cover critical issues,
- The data specification and metadata Implementation Rules, when completed, will be implemented during the design phase of common data format in order to assure their practical usability and to identify obstacles and potential problems,
- To provide guidelines and training.

2.1 Basic cartographic data provided by data providers through WFS (Web Feature Service)

The data are harmonised and converted into one common data model in order to be made available for demonstration, analysis, and access as a homogeneous data base.
To achieve the objectives of the project the following specific actions will be undertaken:

2.2 Design of a common data model

To permit data aggregation in a seamless way, the harmonisation into a common data model is essential. The common data model will be designed and developed based on state-of-the-art specific standards and according to the INSPIRE Implementation Rules. The expert committee established under the project defined a common conceptual model which will include a shared object register, object classification, object definition and an object representation method for chosen thematic fields. Rules will be developed for cross-border data merging, as well as rules for generalisation methodology between four main scales:

- Local level: 1:5.000,
- Regional level: 1:10.000,
- National level: 1:200.000,
- European level: 1:1.000.000.

2.3 Design of data harmonisation process

The next step will be to design a data harmonisation process from a data provider model into the above described common data model. Data harmonisation rules will be produced, in part as general guidelines and in part as individual guidelines for each data provider. Re-modelling of data bases into a common model will be done not by changing the database permanently but by a schema mapping process. On-the-fly harmonisation will be customised for each data provided.

2.4 Aggregation methodology and rules

The aggregation procedures will be implemented when the common data model is in place

2.5 Aggregation process

The aggregation of exposed and harmonised data by the GIS4EU server will be done directly by users using their own tools, procedures and guidelines as provided by the project and by connection with the GIS4EU server or by the created GIS4EU portal, which will provide necessary web services tools.

3. THE MANAGED SPATIAL DATA THEMES

The spatial information, especially at the local and regional level, is difficult to exploit in a European Union context. In fact, the collection and preservation is managed at different governmental levels.

Often collections are not compatible and datasets are not able to be combined with other datasets. For this reason, the project Consortium intent is to collect a wide range of: type, definitions, geographic levels, formats, access rights, use, language and metadata standards.

The project covers and addresses the following major themes referring to the cartographic matter:

3.1 Cartographic themes

The project works on four different spatial themes:

- Administrative units, (INSPIRE ANNEX I),
- Hydrography (INSPIRE ANNEX I),
- Transport network (INSPIRE ANNEX I),
- Elevation model (INSPIRE ANNEX II).

The project manages the previous themes because each of them can focus on different aspects related to the cartographic aspect. For the “administrative unit” aspects like feature definition and cross-border design are some of the first problem to solve to merge the different data sets. For the Hydrography aspects the cross-border definition, the related ontology and the language definition will be an interesting task to manage. All these three cartographic themes are the first level for defining a Geographic European Cartographic Level. About the last theme, the particular designing rules will be a border line to manage 3D aspect at the EU level.

3.2 Cartographic scales

The project will work on four different scales:
- European (1:1.000.000),
- National (1:200.000),
- Regional (1:10.000),
- Local (1:5.000).

The different scales will be used as nominal scales or derived scales, in comparison with the datasets and their standard definitions. The different scales will cover almost all the issue related with this problem.

3.3 Cartographic cross contents

The project will address three different main subjects:
- Cross – Scale,
- Cross – Borders,
- Cross – Languages.

The different cross contents approach would help to address the major issues, comparing them on the three different levels.

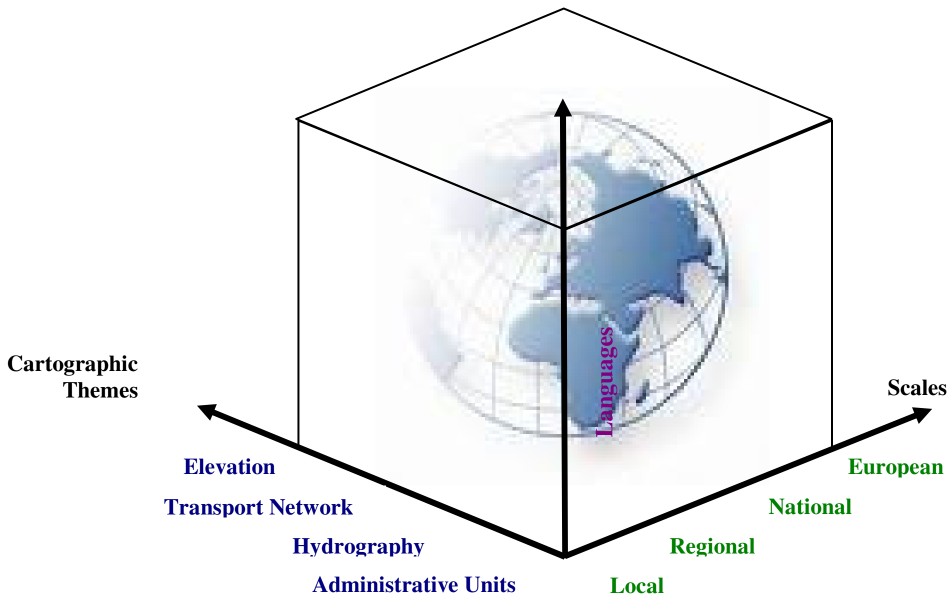

Fig 1. – Logical structure of the project's content

4. THE COOPERATION FRAMEWORK

The project will be developed by 10 different countries, involving 25 project partners. The partners can be mainly classified into the following categories (see Fig x):

- **Data providers** (**DP**), which can be further classified in:
  - national data provider (NDP);
  - regional data provider (RDP);
  - local data provider (LDP);
- **Research partners** (**RES**): including universities, research centres and academic centres;
- **Technological partners** (**TECH**): ICT experts, private ICT companies;
- **Users** (**USER**): representatives mainly associations enclosing a wide number of users.

Each partner will have a detailed role in the project in accordance with the following criteria:

- **DPs** are the owners of the data to be made available in an accessible and usable way. But they are not only the providers of the data, but also and mainly they have the deep knowledge of their data infrastructure and peculiarities. For this reason they play a fundamental role in re-modelling the data structure. The data providers cover 10 countries of the EU.
- **RES**: they are the experts in remodelling design, harmonisation design, data fusion, having strong experience in designing GI data modelling and workflow.
- **TECH**: they are the arm of the research partners: their main aim is to technically support the operational implementation of the rules and the workflows defined by the RES with the critical support of DPs and USER.
- **USER**: they are key actors in the definition of peculiarities of the project deliverables and results. They are the voice of the customers. In order to include the widest number of users, we involve one association representing a very high number of entities acting in GI and GIS in Europe: GISIG, which will collect more than 100 organisations in more then 20 European countries.

The partners, involved in the consortium, have nominated their relevant representatives who will be assigned to the project.

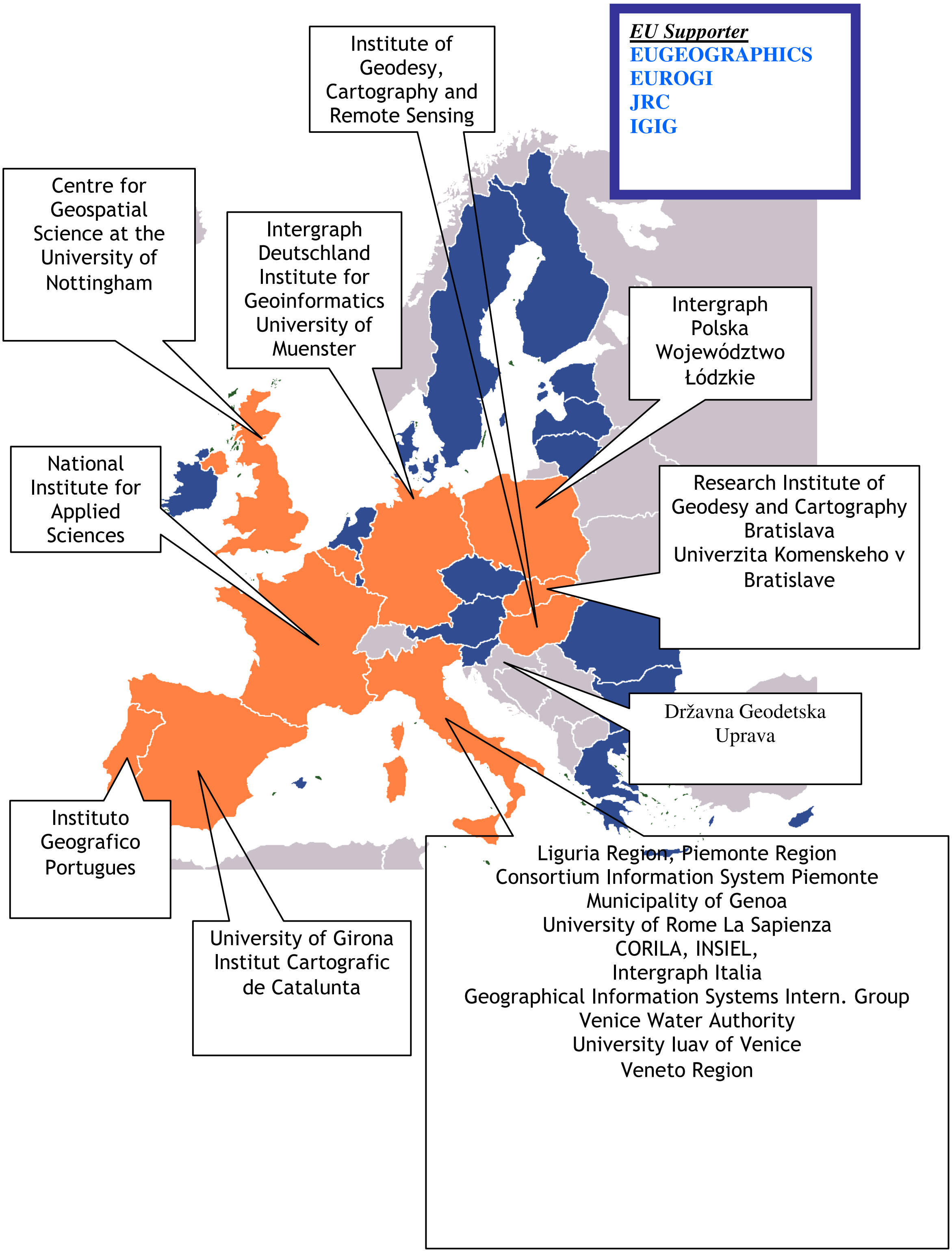

Fig. 2 .– European Countries involved in the project

In addition, we received an official declaration of interest for our project by some of the main important EU players in data harmonisation and interoperability fields.
This acknowledgement highlights the major players from users (EUROGI), to standards (OGC Europe), from INSPIRE directive (JRC-IES-INSPIRE Action) and data providers (EuroGeographics).

- **EUROGI** is the European association of national networks covering the whole European Union and in this project represents the user's perspectives.
- **JRC-IES-INSPIRE** is the JRC SDI Unit and has the task of technically coordinating the INSPIRE initiative and shepherding it through the various steps working towards the realisation of a European Spatial Data Infrastructure. The ESDI will focus on the one hand on the institutional capacity building within the Commission, and on capacity building in the Member States and Candidate Countries.
- **EUROGEOGRAPHICS** is the network of the official national cartographic institutes of the whole European Union and, in this project represents the data providers' perspective.

5. MULTILINGUAL AND/OR MULTICULTURAL ASPECTS

Multicultural and multilingual aspects related to geographic information that represents a barrier in the use/reuse and exchange of data among various levels and different stakeholders involved will be investigated. Such point is very important and the multilingual aspects represented a very important issue, especially in terms of coping with differences from one country to another in the definitions used for the collected data.
In the past the development of SDI focused on the technical interoperability, while problems arising from multilingual and heterogeneous conceptualizations of various user groups did not seem very pressing. Nevertheless in the times of INSPIRE these issues are evidently of particular importance. Because the SDI now face an international perspective, with users from different countries, accessing the same data sources. Providing the SDI with a multilingual support is the key to broader acceptance of SDIs, especially at the local level.
There are several questions that need to be raised and answered before tackling the multilingual task on a larger scale like the INSPIRE initiative. Examples for such questions are the following:

- **What is the aim of the multilingual enablement?**
  Possible answers are: **a)** enabling data and service discovery in a multi-lingual fashion **b)** enabling multi-lingual data portrayal or **c)** enabling multi-lingual data analysis. Naturally more than one aim can be applied for the same SDI.
- **What is the multilingual enablement strategy?**
  Two strategies could be differentiated and during the project we will investigate the pros and cons of each strategy: a) one by one translations of all features for every language involved such that the number of metadata catalogues equals the number of languages. Each catalogue is mapped to the individual datasets b) the unified schema generated by the project is mapped to an ontology that is made multilingual by means of language dictionaries mapped to the ontology
- **How do you technically implement the chosen strategy of multilingual enablement?**

Answering this question requires a complete assessment of the project needs and partner capabilities and requirements. Once the capabilities and requirements are defined, we can build an implementation strategy.

It is obvious that the answers to these questions largely depend on the context. In this project, the main aim is to share the available datasets in an understandable way for the user. This means to provide multi-lingual discovery, retrieval and portrayal.
The multilingual aspect of the project is a multifaceted problem; accordingly it will be approached from different angles. Geospatial information (GI) is often associated with metadata for discovery and indexing. The multilingual problem from a conceptual perspective can be summarized as the problem of retrieving relevant GI whose metadata is indexed with terms which are different from (but conceptually related to) the terms used in the query. The key to a solution then is to build a conceptual framework to which we can map various languages, such that each language terminology refers and maps to concepts that are shared by other languages. Fig 4 summarizes the different layer stack comprising the core of the multilingual problem. The following sections illustrate the various aspects of research that will be conducted within the project to address the problem.

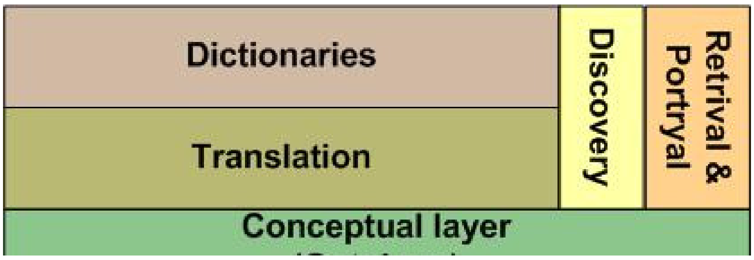

Fig 3. Problem stack of the multilingual aspects for the project

5.1 The Conceptual Layer

Aim of this layer is to solve various problems that might arise when dealing with data discovery in a multilingual environment. To mention some of the problems we consider the following examples:

- *Naming heterogeneities:* An English speaking user might be looking for the term "Highway". This user will fail to find road with the German term "Autobahn" despite the fact it is a correct match. The same user will not benefit from the harmonized European schema simply because his terminology is different.
- *Cognitive heterogeneities:* Within the same language, an Italian user might be looking for the English term "Highway" with a certain conceptualization in mind (how highways are like in Italy. However the query would return all results containing the term "Highway" despite the fact that the Italian user has a specific conceptual description of highways in mind that does not match the English use of the word. Such a problem means the Italian user gets many results but is confused as to which of them maps to his conceptualization.

The conceptual layer intends to solve these problems by establishing a common conceptualization to ensure that equivalent terms from different languages map of the same concepts. Such a conceptual layer might be ontology. Ontologism are defined as a specification of a shared conceptualization.

5.2 Translation and Dictionaries

Language translation is an essential part of the project; techniques for automated translation will be investigated. To make data available in a multilingual fashion (see 5a), the act of translation will probably be limited to keywords for the search in catalogues. The aim 5b) requires more effort while the aim 5c) surely bears the highest complexity and the largest amount of translation work. This includes at the very least the translation of data labels and corresponding legends. Most importantly is the translation of metadata.

We envision the products of translations to be multiple *Dictionaries* for each desired language. Such dictionaries will be mapped to the conceptual layer such that each term in language A has a corresponding concept. Language B is treated in the same manner; this enables querying in any language and receiving the same concepts from the harmonized schema without prior knowledge of the underlying schema in language A or B. Generally many problems would arise during querying such dictionaries, for example we would need morpho-syntactic processing to break composite words in Dutch or German languages. However, this approach is expected to be superior to translation of all available languages into one common language (e.g. English).

5.3 Discovery, Retrieval and portrayal

As illustrated in the translation and dictionaries section, mapping the dictionaries to the conceptual layer will enable seamless discovery of resources in the harmonized schema while masking the various types of cognitive, naming and conceptual heterogeneities from the end user. Furthermore, the data retrieval and portrayal user interfaces have to be translated into the languages of interest. It will be the responsibility of each interested party to provide interfaces in local languages.

However, the project shall provide common specifications for multilingual support for OGC services such as WMS, WFS and their Clients. The multi-language enablement of the OGC services is already part of the OGC specifications which do support internationalization parameters at the service requests given that translated service metadata is available, which can be accessed through the GetCapabilities interface. The Capabilities document is utilized by the Clients to enable interaction with the OGC services and to manage its data layers. Besides the standard service metadata some general service information will have to be provided in multilingual format. Such aspects will need to be further investigated during the project.

To facilitate the access and use of the material in a multilingual context it’s necessary to compare and share the meaning of the definitions used to describe data and metadata collected and maintained by the different data providers. Often we use terms that can be ambiguous without defining the implicit semantics behind them and that are part of the technical and country related culture.

Our layered approach with the conceptual ontology layer enables seamless sharing of the information in a multilingual environment without compromising the consistency of the underlying data model. A public ontology, will be created and discussed under the project.

The adopted solutions consist in the definition of a data model enriched by the semantics to allow easier discovery and searching by using common technological and scientific terms understandable by different kind of user communities operating in the field. A set of core data/metadata which is most used and reused in a large set of different application domains will be identified and semantically modelled it is also possible that common multilingual thesaurus will be defined under the project.
Finally, we propose to use Web 2.0 approaches such as collaborative techniques to enabling building open and evolutionary ontology. Our vision of social semantics (the meaning is in the community) comes into action. Social semantics provides advantages when dealing with large and diverse communities such as the communities of data producers and users in the EU compared with cognitive semantics (the meaning is the minds) and realist semantics (the meaning is in the world). Collaborative approaches to build our conceptual layers will be investigated during the project.

6. EXPECTED RESULTS, SUSTAINABILITY AND IMPACT

6.1 Expected results

Basic geographic information will be aggregated to cover a significant part of Europe; a community of stakeholders will be created with clear and viable plans to take up the results in order to facilitate further re-use and exploitation of the data.
In detail, GIS4EU would like to achieve the following results:

- Common datasets accessible and ***re-usable***
- A cartography sample covering a ***significant portion of Europe***
- A ***common information point*** accessible for all citizens
- Through the information point, data accessible in multiple languages
- A ***transferable*** and ***sustainable*** model disseminated across Europe
- A common multilingual thesaurus
- Synergies with ***INSPIRE*** implementation
- ***Indicators*** monitored and measured

Related with its dissemination activities GIS4EU aims is to motivate new data providers to link themselves to this project.
The advantages for new data providers could be to have at their disposal the operational guidelines developed during the project in order to improve re-modeling data. This involvement will be encouraged by the constitution of a network of data providers depending on class and scale.

6.2 Sustainability and impact of the project

The sustainability of the GIS4EU project hinges on the assumption that it will have a long term impact on its partners, users and stakeholders. As far as the project is concerned, this purpose will be ensured by the following conditions.

6.2.1 Consortium conditions

Thanks to the partnership composition of GIS4EU - including both commercial users and content providers - expected viability will be built on project results.
As far as **content providers**' financial sustainability is concerned, the consortium is largely composed of public sector bodies (national cartographic institutes, regional and local public authorities) that are considered financially viable to ensure:

- Full conversion of data into new data model resulting from the project,
- Exploitation of the project's guidelines to solve cross-border, cross-scale, and cross language problems in other context.

6.2.2 External conditions

Thanks to the development of the planned dissemination activity GIS4EU will be able to exploit results beyond the project.
This will be ensured by the use of different dissemination channels (academic, public and commercial network) that will be able to attract relevant stakeholders, active at Community level, and therefore contributing to the following areas:

- Thanks to the growing interest at Community level in **cross-border policy areas**, GIS4EU is expected to have a strong impact in supporting new application in the field of transportation, agriculture, disaster and risk prevention and management, and in mobile communications services.
- As far as future development related to **European Policy and Initiatives**, GIS4EU will be useful for a great number of applications, mainly in environmental and security fields, such as GALILEO, GMES, Single European Sky, INSPIRE.

6.3 Knowledge and awareness as main project results

At the end of the project, different web products, like web site with a web portal, allow exposing the results and the deliverable products as reports and guidelines, will be left available on the web.
All this material will be available after the end of the project and freely downloaded from the web site developed.
In particular, what developed will be kept available throughout some of the consortium partners:

- CORILA as project coordinator;
- INGR as technical support and for technical and technological aspects;
- CSI and ICC as data providers to make available the data set to expose as demo

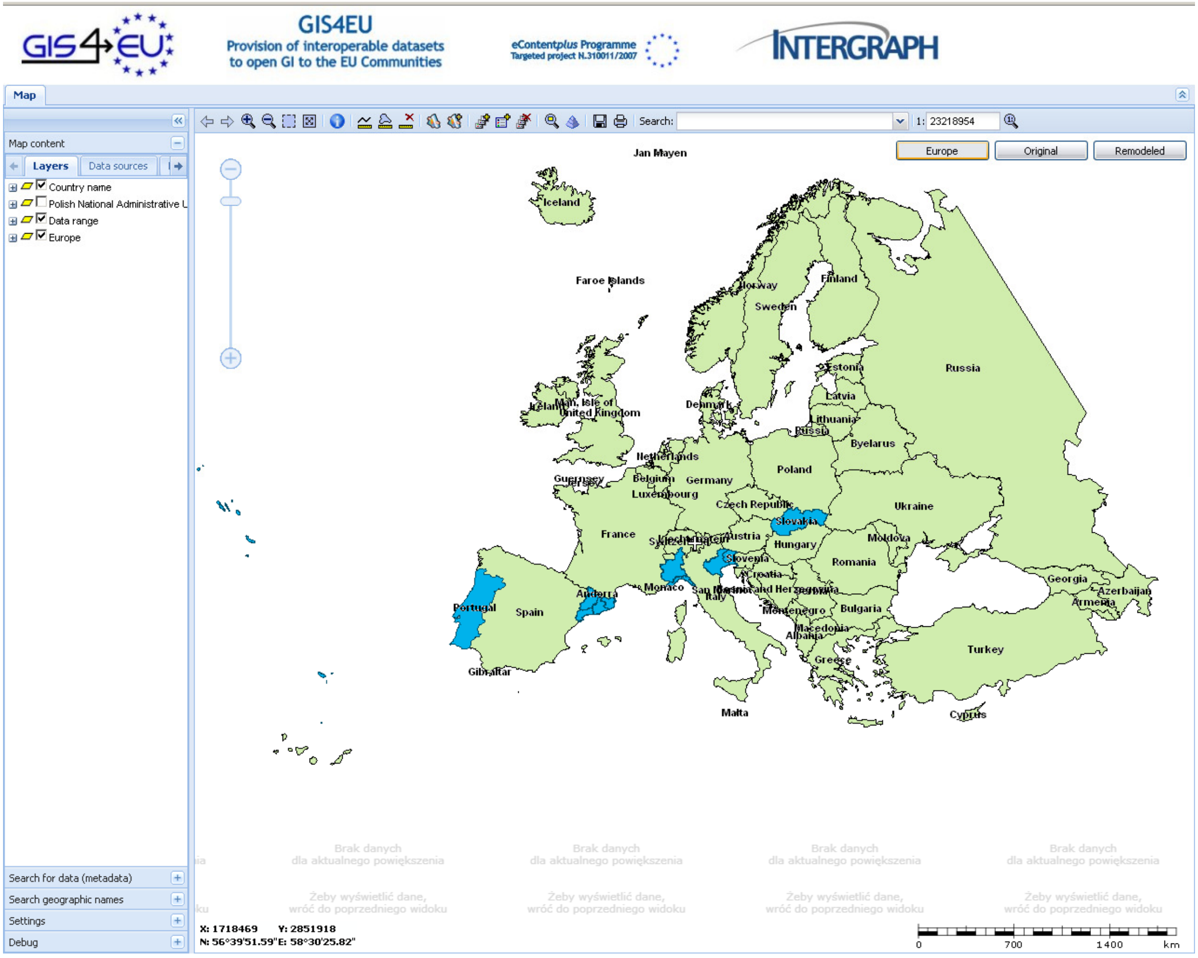

Fig 4. The prototype thematic geoportal

This will allow to keep available the project results and awareness also after the end of the project period.
All data providers have agreed to leave a demo data set, a sub set of project data, available free of charge on the web portal, to allow that all services and data exposition methods developed during the project can be used after the end of the project activities.
To make available to the stakeholders' community what is being developed during the project, the web portal developer will be linked to the JRC **EU Geoportal.** This will be also linked to the objective of increasing the exploitability and visibility of the project results.

6.3.1 <u>Network enlargement</u>

To increase the impact of the project stakeholders, enlargement activities will be developed with some of the project financial resources.
The involvement of other stakeholders could represent different situation and problems about data set and allow increasing the European coverage of GI data sets.

Within the dissemination and awareness raising work package, network promotion material will be developed. Envisaged conferences and workshops will be used to make contacts to the relevant GI stakeholders.
Some of the most important European consortiums, such as EUROGI, IGS, AGILE, AGISEE, GISIG, are already successfully involved, supporting the project and the subjects addressed by GIS4EU.

6.3.2 Collaboration with JRC

To maximize benefits of INSPIRE, GMES and GALILEO, GIS4EU consortium will collaborate and exchange information with Joint Research Centre (JRC) as the technical responsible for INSPIRE.
This will support the project to build an integrated view of experts and to create integrated guidelines, standards, and to implement best practices during the time period in which the technical specification of INSPIRE are going to be create.

ACKNOWLEDGEMENTS

The authors have expressed their thanks getting the contribution from eContentPlus 2006 programme under contract No#GAN310011.

REFERENCES

Project URL: www.gis4eu.eu
Results dissemination international workshop: www.inspire.gis4eu.eu/
Thematic workshop on 3rd INSPIRE Conference 2009:
www.gsdi.org/gsdiconf/gsdi11/workshops.html#wrkshp2.3

BIOGRAPHICAL NOTES

Dr Szabolcs Mihály, Director General of "Institute of Geodesy, Cartography and Remote Sensing (FÖMI)". Membership of professional bodies :Consulting member of Geoscience Department of Hungarian Academy of Sciences (HAS); Deputy chairman of the Geodetic Scientific Committee of HAS; Chairman of Geoinformation Standard WG of the Hungarian Body of Standards; President of the Hungarian Surveying, Mapping and Remote Sensing Society; Hungarian delegate to Commission 3 of the FIG (International Federation of Surveyors); Hungarian representative to Eurogeographics, which is Association of the National Mapping and Cadastre Agencies (NMCA) of European Countries. Teaching experience: Lecturing at Department of Geoinformaatics of the University of West Hungary, Székesfehérvár, associate professor on GPS, GIS, digital photogrammetry and digital cadastre surveying, 1989 – to present. Periodical lecturing at the Budapest Technical University, honorary associate professor on GPS and Land Information Systems, 1985 – to present.

Pál Lévai, graduated as MSc in Geodesy at the Technical University of Budapest, Faculty of Civil Engineering. Work includes: practice on geodesy and related R+D field; development of astrolabe measurement system, establishment of the Väisälä interpherometric baseline at Gödöllö, Hungary; introduction of GPS technology in Hungary; network densification; deformation measurement. Following the major activities fields working in the INSPIRE team at FÖMI; practice on GIS; spatial data management and EU project management field. He is board member of Geographical Information System International Group (GISIG, Genoa, Italy) and member of the Hungarian Surveying, Mapping and Remote Sensing Society.

CONTACTS

Dr. Szabolcs Mihály
Institute of Geodesy, Cartography and Remote Sensing
H-1149 Bosnyák tér 5.
Budapest
HUNGARY
Tel. +36 1 222 4001
Fax +36 1 222 4012
Email: mihaly.szabolcs@fomi.hu
Web site: www.fomi.hu

Mr. Pál Lévai
Institute of Geodesy, Cartography and Remote Sensing
H-1149 Bosnyák tér 5.
Budapest
HUNGARY
Tel. +36 1 460 4253
Fax +36 1 222 5118
Email: levai.pal@fomi.hu
Web site: www.fomi.hu