Incremental Construction of a Regional SDI, an Example Case in the Galicia Region

Rubén Béjar[1], Jacobo Aboal[3], Michael Gould[2], P.R. Muro-Medrano[1], Pedro Vila[3]

[1]Department of Computer Science and System Engineering, University of Zaragoza
María de Luna 1, 50018 Zaragoza (Spain)
rbejar@unizar.es, prmuro@unizar.es

[2]Department of Information Systems, University Jaume I
Campus de Riu Sec, 12071 Castellón (Spain)
gould@lsi.uji.es

[3]Consellería de Medio Ambiente - Xunta de Galicia
Rúa San Lázaro, s/n Santiago de Compostela 15.781, A Coruña (Spain)
jacobo.aboal.vinas@xunta.es, pedro.vila@xunta.es

Introduction

Public administrations have been facing problems with geographical information use and management for years: difficult software interoperability, incompatible data formats and little knowledge about the available geodata are among the most common ones. This was the case of the Galicia Department of Environment (Consellería de Medio Ambiente, Xunta de Galicia (CMA)), in charge of managing environmental resources in that region. INSPIRE work in progress toward a European Spatial Data Infrastructure [JRC], currently focused on environmental issues, led to a solution based on this European Commission initiative recommendations, building thus the core of a Galician SDI.

Galicia is a NUTS-2 region located at the northwest corner of the Iberian Peninsula. The climate is warm and wet so its land is covered with many forests (69% of its surface). This fact makes forests the main concern of the CMA, with water use, disposal of waste and protected natural environments among its other responsibilities.

The decentralization of the CMA, nineteen districts, four provinces and the central office in Santiago de Compostela, makes the usual problems with geographic information in big organizations and public administrations much worse. It's difficult for users to find the data they need or even to find out if that information exists. In some cases, i.e. people in forest districts away from the central building, users just didn't have any access to geographic information that would make their work much easier. Another problem is the delay in the building of an integrated solution that has led some districts to adopt different GIS software solutions, or no solutions at all in some of them. This situation has added problems of data formats and interoperability.

The situation presented a challenging task, because almost everything needed to be done in order to provide CMA users with GIS information and tools. Given this, and following INSPIRE recommendations, the development of this SDI was planned in an incremental way: several phases are planned and every phase produces some infrastructure services and some user applications. This strategy makes the infrastructure useful and visible from the beginning,

making it easier to get funding to continue with its development, and allows for giving real GIS services and applications to the CMA users.

The paper will show the planned phases, the user applications that are being developed as the infrastructure grows and will give an overview of its current state of development (first phase finished, second phase in progress). This strategy of incremental development may be a pattern to be followed by other regional governments starting to develop their own SDI.

Incremental Construction of SDI

Spatial Data Infrastructures (SDI) are a collection of technologies, policies, institutions and people that facilitate the availability and access to spatial data [GSDI].

From a technologic point of view they are distributed, potentially complex, geographic information systems based in evolving standards. On one hand their complexity leads to long development times and the fact they are based on evolving standards leads to the necessity to address continual change. On the other hand, long term projects are difficult to sell to public administrations, where budget needs usually to be justified year by year, and more when the benefits of these projects are not expected in the short term, as may be the case with SDI. In the rest of this section an incremental, iterative, approach to the development of SDI is proposed, as the way to address this kind of problems.

Iterative Development for SDI

Iterative development is well-known way to develop software projects that aims to show quick results, even incomplete, allowing thus for early inputs from the users in order to adapt better the system to their necessities. A side effect of this approach is that final users see results quickly so their confidence in the success of the project is easier to keep.

Iterative development methodologies usually include most common phases in software projects at every iteration (i.e. and very briefly: analysis, design, implementation and testing). The purpose of this paper is to try to set up some guidelines on approaching SDI projects making profit from the benefits provided by iterative development though from a different scale: the point is not only applying iterative development to the building of SDI, but extending this idea to the long-term planning of this kind of projects, having 4 to 6 months length phases, iterations, with concrete and useful results at the end of every phase. The proposed main steps in every one of these phases should deal with:

1. User's requirements: SDI has a clearly defined purpose, so capturing user's requirements seems not so important beyond the first phase. But in order to sell them to the decision makers, specific benefits for their organizations or administrations must be stressed. The idea is building SDI to solve problems and facilitate daily work in these organizations so funding is easier to get and justify. Other important point here is the necessity to collect new user's requirements for every phase. This kind of long term projects need the continual support from the decision makers to assure the funding needed to complete them, and this support needs from evidence of continuous improvement and growth.
2. Spatial data: spatial data need to be created or obtained, modified, updated, corrected etc. and this is a continuous process that needs to be addressed periodically to keep

quality standards and to give access to this new or updated data to the users (i.e. making it available through standard web services). Every phase should review the existing data or new data necessities that may have arisen and incorporate new or modified data to the SDI.

3. Metadata: metadata evolve as data evolve, though most metadata are usually more stable than the data they describe. Again every phase in the development of SDI should be used to review the existing metadata, though its maintenance should be a continuous process.
4. Web based services: spatial data infrastructures are based on standard web services. At this moment, there are some web services specifications well established but there are many others only recently published or even not finished. Besides this, setting up and configuring these standard web services (i.e. adapting data to be served by them) is usually a complex task, so deciding carefully which services deploy and which data must be behind them is important to show early results and not having too long development times.
5. User applications: final applications, the only part that many users will see from SDI, must guide the development of every phase. This way, both users and decision makers will see useful results at the end of each phase, what will make it easier to keep their confidence in the project. Deciding which applications can be developed to make the bigger profit from the services developed in every phase is thus a cornerstone of this incremental building strategy.

An example case in Galicia: the IDE Rimax project

Given the problems related to geographic information management and use in the CMA, and the development of the INSPIRE initiative with the requirements it will impose to EU members in some years, building an SDI following this initiative principles was the best option to address both issues simultaneously. This would solve the CMA geographic information users' needs while giving some effective steps in order to fulfill the future INSPIRE legislation, making profit of the recent networking of all delegations and districts. As more information on the SDI being built for the CMA (the IDE Rimax project) can be accessed in [IDE Rimax], and a more detailed technical view in [Béjar R., et al], the rest of this section will focus on describing the incremental approach taken to build this SDI, as a concrete example of the ideas presented in the previous chapter.

Iterative development for the IDE Rimax project

The CMA presented several necessities that were not easy to address together at the beginning of this project:

- Early results: all CMA districts were being finally networked, after a strong inversion given the big decentralization of this department, and it was important to start exploiting this infrastructure by providing services to their users.
- Standards compliance: the people responsible for the IDE Rimax project were convinced of the necessity of building their GIS on standards, given their advantages (i.e. vendor independence).

- Building a spatial data infrastructure: the INSPIRE project made it important to design a system able to become the core of a future Galicia SDI. This involved addressing some problems (i.e. a stricter partitioning on standards-based services, or a more detailed analysis on models) that were note urgent and thus difficult to "sell" to the decision makers.

The necessity to build an SDI based on the INSPIRE proposals was more or less evident, but the way to combine this with the necessity to have early, and final, results was not so obvious. An incremental approach was designed as the way to overcome these problems. At the moment of writing this paper, the first phase is over, and the second phase is advancing, so more details can be given of them. A hypothetical, depending on budget etc., third phase is for now in an early stage of planning. An overview of the planned phases, with a reference to the steps described in the previous section, is presented next:

Phase 1 (Finished)

1. User's requirements:
    a. all geographic information must be stored in a spatial database
    b. it will be possible to find relevant geographic information available
    c. users will be given tools to view the geographical information they need over the CMA Intranet. Users will also be given the possibility to download the geographical data (vector and raster coverages) in the format they are used to work in
    d. advanced users will be able to access simple map services, designed to be combinable with their local data
    e. software already owned by the CMA will be used where possible
    f. standards must be followed, if possible, where available; specifically those standards recommended by INSPIRE and GSDI
    g. architecture will follow that recommended by the INSPIRE project work groups
2. Spatial data: the objective in this phase is loading the most needed data for the CMA users, so covering most of Galicia at the largest possible scale is looked for
    a. spatial database (Oracle 8i) installation and configuration. In this phase exploiting the spatial capabilities of the database is not considered but to a minimum because of the lacking of complex data models
    b. large scale topographic information (agreement with public works department; includes administrative boundaries, communication infrastructures, public services etc.). This is targeted to forest rangers, so large scale, precise, information is required
    c. forests (on of the most important concerns in the CMA and maybe the most important information they create and maintain)
    d. protected natural areas (mainly natural parks)

   e. orthoimages (low to medium resolution imagery covering most of Galicia region and some high resolution images of natural parks)

3. Metadata: the objective is providing basic metadata for all the information loaded into the system, in order to get users familiar with them, in a short place. These metadata should also be a good basis for a more detailed metadata creation process
   a. Minimum (based in the Dublin core metadata initiative) metadata are created for the data loaded into the IDE Rimax
4. Web based services: the basic services in SDI are catalog and web map services, so in the CMA there were installed:
   a. The Java and Oracle based catalog service developed by our laboratory.
   b. ArcIMS was used to provide web mapping capabilities, thanks to its OGC connector that provides the web mapping standard in its version 1.0. Services were created for the core and environmental data detailed in point 2.
   c. A partial web coverage service wrapper was developed over ArcSDE to provide standard raster data download capabilities.
5. User applications: essential to provide functionality to end-users, the following were developed over the created web services:
   a. Basic search engine, in order to familiarize users with the catalog and to provide a simple point of access to it.
   b. Forest rangers thematic visualization application, designed to provide customized visualization capabilities to forest rangers with the data they need for their daily work (some very detailed topographic information, and environmental data).
   c. Basic, generic, visualization and data download applications, to provide simple access to the map service and the partial coverage service developed.

Phase 2 (Ongoing)

1. User's requirements:
   a. users will be given the possibility to make spatial and non-spatial queries to the core geographic information through standard web feature services, though with basic data models
   b. metadata will be in different languages (Spanish, Galician and English). The searches and applications making use of these metadata will be able to exploit them in the most appropriate language
   c. public environmental geographic information will be made public through the appropriate standards
   d. tools will be provided to access seamlessly to different standard map services over the Internet. This will include changes of projection, coordinate transformations etc. where needed

   e. a basic Galicia gazetteer will be implemented based on standard web gazetteer service technology
   f. there will be basic mechanisms for forest rangers to submit updates in the areas they are responsible for

2. Spatial data: this phase completes the loading of the basic information for the CMA users, but it is also aiming at providing more general spatial information on Galicia in the line of SDI
   a. high resolution imagery of the main cities in Galicia
   b. Natura 2000 information
   c. medium scale topographic information (most appropriate for general users). This will become the source of the geographic names for the gazetteer
   d. procedures will be defined to incorporate updates coming from the field agents to the spatial database
3. Metadata: on one hand, this phase must consolidate the metadata creation process started in the previous phase. On the other hand a process to translate the originally created metadata, in Spanish, to English (to facilitate collaboration from the IDE Rimax with other European SDI initiatives) and Galician must be addressed
   a. Creating Dublin core metadata for the new loaded information
   b. Improving the metadata for the most important information created and maintained in the CMA (forests), taking compulsory ISO 19115 fields as a guide
   c. Translating the metadata to English and Galician. This process will be facilitated by the tool chosen for metadata creation (CatMDEdit, developed by our laboratory).
4. Web based services: to consolidate the basic services installed in the previous phase and to install some new ones (feature service and gazetteer) for some simple new capabilities.
   a. Multilanguage-enabled catalog service.
   b. New web map services for the new data.
   c. Web feature services for basic data (administrative boundaries, environmental information…)
   d. Web gazetteer service, based in medium scale topographic maps information.
5. User applications: to make profit from the new services providing more end-user utilities.
   a. Improved search engine, incorporating multi-language search capabilities.
   b. Natura 2000 information public visualization and querying.
   c. Geographical names search and visualization (to make profit from the gazetteer service).

d. Forest exploitation management application.

Phase 3 (Early planning, only some ideas)

1. User's requirements (to be refined):
    a. developing advanced data models for the most important datasets created and maintained in the CMA
2. Spatial data: incorporating new data created or obtained. Preparing procedures to update existing information.
3. Metadata: updating metadata as needed.
4. Web based services:
    a. New web feature services to make profit from the data models.
    b. Distributed catalog services.
5. User applications: applications that can provide functionality based on the new data models.

This phase would also be a good moment to prepare some pilot projects to connect this SDI with other SDI initiatives in the Galicia region, to test interoperability and establish some common, needed, foundations.

Conclusions

The incremental approach described in this paper has allowed for having visible, useful results every 6 months or less in the development of the spatial data infrastructure for the CMA. This has provided a good relationship with the users, who were able to offer their inputs and suggestions from the beginning. This approach also has facilitated the construction of new applications and services naturally built on top of the existing ones, making profit from the web services architecture proposed by the INSPIRE initiative. Long processes, like creating complete metadata, have been naturally divided and addressed in steps, thus allowing for making profit from them from the beginning, even if they were not completed. Finally, training users has also being easier because the new concepts and applications are being presented one by one, so users are given enough time to understand them. The IDE Rimax is now solving real CMA users necessities while still being able to become a full member of INSPIRE, because it is being developed following the appropriate architecture and standards suggested by this initiative.

Acknowledgments

The basic technology of this work has been partially supported by the Spanish Ministry of Science and Technology through the projects TIC2000-1568-C03-01 and FIT-150500-2002-347, and by the project P089/2001 of the Council of I+D of the Government of Aragón. The specific design and implementation reported here was sponsored by two R+D contracts to the authors, by the Xunta de Galicia, Conselleria de Medio Ambiente (CMA).

References

Béjar, R., Aboal, J., Gould, M., Muro-Medrano, P.R., Vila, P. "Spatial Data Infrastructure in the Galicia Region: An Instantiation of the INSPIRE framework", Proceedings of the 6th AGILE Confrence, pp. 267-277, April 2003.

GSDI. Global Spatial Data Infraestructure homepage. [Online] http://www.gsdi.org

IDE Rimax. IDE Rimax information homepage [Online] http://inspire.xunta.es

JRC. JRC-Institute for Environment and Sustainability, Ispra (publisher). "INSPIRE Architecture and Standards Position Paper". [Online] http://inspire.jrc.it/reports/position_papers/inspire_ast_pp_v4_2_en.pdf, 2002.