CO-410

BREAKING DOWN BARRIERS BETWEEN CARTOGRAPHY, GEOSPATIAL INFORMATION AND ENVIRONMENTAL MONITORING DATA: TOWARDS A NEW PRODUCTION MODEL

VALCARCEL N., AROZARENA A., VILLA G.

National Geographic Institute of Spain, MADRID, SPAIN

ABSTRACT

Environmental Geospatial information is required by several International and European and National agreements, commitments and legislation such as the Framework Convention on Climate Change of the United Nations, Kyoto Protocol, Habitat and Water Framework Directives of the European Union, etc. This specific category of cartographic data has a key requirement that should be considered as essential to set up an efficient production system: must be free and open information for European citizens, as was established in the 1998 Aarhus Convention, ratified by Spain on 15 December 2004, which has represented a milestone in the advancement and development of public participation in the conservation and defence of the environment. This key requirement is also clearly reflected in the European Inspire Directive.

However, when contemplating the cartographic production of environmental information, both in terms of thematic content and geometric accuracy, there are important gaps in what the users of environmental information get and really need (managers and analysts), when geospatial databases and inventories were obtained following a traditional production model in the collection and processing of cartographic data.

Thematic contents and data modelling techniques used by data producers have had a great importance on the quality and usefulness of the data available for environmental monitoring and analysis, such as the use of rigid hierarchical classifications when categorizing and grouping information. In addition, purely cartographic aspects such as minimum mapping unit and scale affect strongly the resulting statistical reports of geospatial environmental inventories. One of the best examples appears when comparing land cover statistics of European Corine Land Cover Databases (1:100.000 scale, with Corine Land Cover Level 3 nomenclature) with National Land Cover inventories. The same problems will appear when comparing this National land information with GMES High Resolution Layers, this time in geometric raster or grid data structure, which eventually could cause disparity when analysing National and GMES data for Greenhouse Gasses emissions and removals in the LULUCF sector.

Therefore, and especially taking into account the INSPIRE guidelines, it is necessary to modify and improve the whole production flow of environmental geoinformation, through dialogue and active cooperation of the main stakeholders and users of such information (European and Global user of environmental data, National Mapping Agencies, Environmental Ministries and Agencies and other related public and private corporations). There are three pillars for this cooperation:

1) Data Model able to respond to real user needs and requirements, designed by consensus between stakeholders. This in practice can be reached using the standardized modelling techniques (ex. UML, Object Oriented design) specified by the ISO TC 211 standards and Inspire specifications and recommendations.

2) Use of cost-efficient normalized land monitoring data, mainly imagery (satellite images and ortophotography), as an objective reference data for extracting geospatial information.

3) Flexible production flow, in a double bottom-up and top-down approach, for harmonizing technical and social requirements with real social and budgetary constrains, taking into account the ‘free and open’ characteristic of the environmental geodata.

1. BACKGROUND AND OBJECTIVES

The Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters, known as Aarhus Convention, was adopted on 25th June 1998 in the Danish city of Aarhus at the Fourth Ministerial Conference in the 'Environment for Europe' process. In its article 4, ‘Access to environmental information’, says ‘Each Party shall ensure that, subject to the following paragraphs of this article, public authorities, in response to a request for environmental information, make such information available to the public, within the framework of national legislation’. This obligation lead to the entering in force of the INSPIRE Directive in May 2007, establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment. Other international agreements, such as the Framework Convention on Climate Change of the United Nations, Kyoto Protocol and the Convention on Wetlands or “Ramsar Convention”, highlight the importance of geospatial information for environmental monitoring at the international and regional (European) context. The strong obligation established by Aarhus Convention and later by Inspire regarding the open access to environmental geospatial data leads to a reorganization of the public environmental cartographic datasets. This reorganization, guaranteeing the open access to the information, should consider the whole production chain between public and private stakeholders to make it sustainable on time and funding.

In terms of cartography, environmental mapping covers how to integrate a diverse set of spatial data and maps that may include environment-related variables such as land cover and land use, soil types, vegetation, water quality, climate, geochemistry, and even the geology and geomorphology (Parry, 1989). This type of mapping therefore makes partitions of geographical space in form of units that indicate one or more environmental variables.

The development of this type of mapping requires integration of both capture techniques and production techniques of cartographic data, but also data collection and representation directly or indirectly representing variables or indicators of the environmental field. With the development of information technology in the last 25 years there has been much easier and cheaper generate information in both types of production, which in practice meant an exponential increase in the volume of environmental cartographic products. Precisely this ease of capture and production of thematic geographic data has led, in many cases, a multiplicity of inventory and thematic map databases which have been developed by different institutions and both public and private corporations, in most cases representing the same objects or geographical spaces, but just looking some specific environmental aspect or feature thereof. That is, there are often multiple representations of the same occurrence or geographical phenomenon, which results in a duplication or multiplication of its geometric representation in the various environmental inventories. The user is thus faced with the problem of choosing one geometrical representation of the phenomenon, in many cases when is using several environmental maps or datasets simultaneously.

Second, many environmental inventories began to be produced (about 30 year ago) precisely for cartographic purposes only, including analogical data capture (using stereoscopic photo interpretation), so that the storage of information was designed according to the possibilities of representing data on a paper map. If we add the technical limitations in the emerging geographic information systems, we find time series of environmental maps printed on paper or in layers in the digital case, which many times summarized groups of biophysical and environmental variables as labels. That is, the use of nomenclatures, classifications and thematic map legends directs the generation of the first environmental geographic inventories, and those classifications, conceived with hierarchical modelling, makes very difficult to query those inventories outside the initial requirements, and therefore, out of the initial purpose for which they were conceived.

But looking at again to International Conventions, International Reporting obligations undertaken by countries actually pose strong requirements in terms of very detailed environmental spatial information, and very frequent update cycles. A key example is the IPCC Guidelines to elaborate the projections of GHG emissions and removals in the LULUCF sector, for the National Greenhouse Gas Inventories. There have been considerable amount of changes in reporting categories, structure of the reporting, new source categories and revised methodologies in the 2006 IPCC Guidelines require a recalculation of the time series of GHG emissions to avoid time series inconsistencies due to the introduction of the 2006 IPCC Guidelines. Some of the problem related with the use of hierarchical nomenclatures arise when evaluating land areas in six general land cover categories (Forest land, Cropland, Grassland, Wetlands, Settlements and Other land). Other categories of problem are related with the equivalent cartographic scale of the different National Inventories, geometric and thematic accuracy, and uncertainty. The third problem arises with the need to take periodic inventories, referred to fixed dates in time, which does not usually coincide with the long update cycle (5 to 10 years) of National Thematic Inventories (Forest Map, Agricultural Maps, etc.).

2. APPROACHES

2.1 The European experience

One of the best examples of environmental geodatabase designed and produced in a traditional way is the European Corine Land Cover (CLC) and Land Use series, coordinated by the European Environmental Agency (EEA). Although Corine Land Cover meant a very important first step in terms of standardization and harmonization of information in Europe, is still strongly conditioned by its own hierarchical data model. Among the most important problems of its hierarchical nomenclature are:

- Proliferation of unnecessary "threshold subclasses", due to differences introduced by values adopted by a particular parameter or to the need to address some of the situations in which a polygon contains several lu/lc types, and providing little information to the user which do not know in depth the definition of the class

Complexity of class definitions makes it difficult to build, update and use the databases.

- Information stored in the CLC database is much less than information acquired by the photointerpreter. Important spatial variations in certain parameter values do not appear in the database. Polygons with very different percentage of buildings, or trees,.. cannot be differentiated in the information system if this variations do not "cross" the threshold line, and is not possible to make calculation of parameters and Indicators based on parameter values.

- Important temporal variations in certain parameters can not be obtained of the CLC Change database, because these variations do not "cross" the "definition rule" threshold and/or these variations are hidden in polygons assigned to dominant classes or to mixed classes.

- It is not possible generalization or transformation of the CLC database out of its own hierarchical system, due to the loss of information regarding those key parameters used to define CLC classes (% building or sealed soil, % tree crown density, etc.).

The second problem concerning CLC is the Minimum Mapping Unit (MMU) concept, to be applied when producing cartographic dataset with polygonal geometries, for obtaining a usable and cost-efficient information according to requirements. For CLC MMU is 25 ha., and CLC Changes applies a MMU of 5 ha. This operative unit for capturing surface information becomes a problem when is combined with a rigid hierarchical nomenclature, because forces the photointerpreter, in many cases, to draw polygons that contain areas with different land cover. In spite of this, he must assign the polygon to one, and only one, class of the Nomenclature. MMU established by European needs (approximate cartographic equivalent scale of 1:100.000 for CLC) had two consequences in the CLC Nomenclature:

a) Need for "flexible" class definitions: CLC assign the polygon to the "dominant" class among those present in the polygon: normally the class that occupies the greatest percentage of the polygon's surface).

b) Need for Mixed classes: in cases in which there is no clear dominance of any class, there is a need for including some "mixed classes" in the Nomenclature

Since late 80's the EEA has coordinated the production and later use and divulgation of the CLC series in Europe as open access data for all European users, applying in practice the principles of Aarhus, even before they were established.

But the increasing demand of environmental monitoring for Europe lead to three different initiatives in the XXI century, in order to provide enough and open environmental spatial information:

• Infrastructure for Spatial Information in the European Community (INSPIRE): The entering in force of the INSPIRE Directive in May 2007 established an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment. INSPIRE is based on the infrastructures for spatial information established and operated by the 27 Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. One of the strengths of INSPIRE is the definition of the Data Specifications of these 34 themes, so that there is an effective data harmonization in the form of data models for sharing spatial information on environmental data. For the definition of these Data Specification, group of thematic expert have been formed, and will be working till 2012.

• Shared Environmental System (SEIS): SEIS is a collaborative initiative of the European Commission, the EEA and its member countries, which aims to improve the availability and quality of information needed to design and implement the European Union's environment policy, streamline data handling by connecting existing information systems and providing online information services, and modernize environmental reporting to reduce the administrative burden both at national and international level. SEIS then focus on availability and reporting of information, while Inspire deals with data harmonization and sharing, always using the best official information available, which, in most of cases, is produced with high thematic detail by countries, Inspire and SEIS then provide the adequate platform for data sharing at pan-National and Regional (in this case, European) level.

• Global Monitoring for Environment and Security (GMES): GMES consists in a complex set of systems which collects data from multiple sources (earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors), processes these data and provides users with reliable and up-to-date information through the services mentioned above. Six thematic areas are to be developed: marine, land, atmosphere, emergency, security and climate change. A land monitoring service, a marine monitoring service and an atmosphere monitoring service contribute directly to the monitoring of climate change and to the assessment of mitigation and adaptation policies. Two additional GMES services address respectively emergency response (e.g. floods, fires, technological accidents, humanitarian aid) and security-related aspects (e.g. maritime surveillance, border control). The coordination and management of the GMES programme is ensured by the European Commission. The setting up of initial versions of the GMES services have been assigned to several projects partly financed through the 7th Research and Development Framework Programme of the European Union, while the developments related to the observation infrastructure are performed under the aegis of the European Space Agency for the space component (i.e. Sentinel missions) and of the European Environment Agency and the Member States for the in situ component. Considered as "public goods", GMES services should be accessible to any organisation or citizen.

Inspire and SEIS are working in the same line for harmonized the existing and future environmental spatial data in Europe, but GMES first technical proposal was conceived as a Top Down program, providing centralized open land, marine, emergency and atmospheric data services for European users. This first design is now be improved toward a flexible Top down+ Bottom Up approach, in order to be producing taking into account the National inventories, according to the Inspire Directive. This flexible approach is being now considered during GMES Initial Operation in the 2011-2013 period.

2.2 The Spanish experience

In Spain Law 27/2006, of July 18, established the rights of access to information, public participation, and access to the judiciary for environment related issues (incorporating European Directives 2003/4/EC and 2003/35/EC). Concerning cartography, two laws applied this principle to the Spanish cartographic production:

- Law 14/2010 of 5 July, on infrastructure and geographic information services in Spain.
- Royal Decree 1545/2007 of 23 November, which regulates the National Cartographic System.

This legal framework, led in the cartographic level by the National Geographic Institute of Spain (IGN), ensures the open access to essential environmental geodata such as orthoimagery, land cover &use data, hydrographical spatial inventories, etc. But there is still a need for coordinate environmental cartographic production in Spanish Ministries and Autonomous Regions, according to the administrative decentralization of the nation. Some important steps have been given by the National Cartographic System, and more specifically, by the National Land Monitoring Program. Further coordination should be done by Central and Autonomous Administrations, in order to avoid duplication in production of thematic geodata such as forest inventories, agricultural maps, etc. SIOSE (see above) project is a good example of how, applying modern techniques in the design and production of spatial databases, different public stakeholders can coordinate the production of essential environmental geospatial data (land cover), avoiding duplication of data capture, and producing a unique inventory which can be analysed with multiple criteria by different users (environmental, agricultural, forest, urban planning, etc.).

2.3 PNOT: National Land Monitoring Program - SIOSE: National Land Cover & Land Use Information System

IGN, according to its role as National Mapping Agency, is coordinating since 2005 the PNOT. In its first phase, ‘Capture and processing of aerial and satellite imagery’, PNOT obtains aerial and satellite coverages with resolutions and updating periods cost/benefit optimized for the applications in which they would be used. These coverages are organised in different levels of spatial resolution (pixel size) and temporal resolution. This phase is performed through two National Programs:

- National Remote Sensing Program (PNT): provides regular coverage of Spanish territory with satellite imagery, both present and historical.
- National Program for Aerial Orthophotography (PNOA): provides regular coverages (each 2/3 years) of the whole Spanish territory with digital aerial imagery, making it possible the massive use of this data.

In PNOT’s second phase, SIOSE, Land Cover and Use Information System of Spain, integrating different land cover inventories of regional and national administrations, is managed and coordinated by the IGN, and produced in a decentralized and cooperative model by Spanish National and Regional administrations. This projects was designed according to the main INSPIRE principles, and ISO TC/211 standards, moving from previous hierarchical land cover databases towards a land cover feature data model, able to describe the different land cover classes in the territory, but also their environmental parameters. IGN, in its role as National Reference Centre in Land Cover and Use, manages and coordinates SIOSE during its first production (2006-2009) and its updates for temporal reference 2009 and 2011.

2.3.1 SIOSE Technical Features

• Geodetic Reference System: ETRS89, according to INSPIRE and Spanish Geographical High Council.

• Cartographic Projection:UTM, zones 28, 29, 30 y 31
• Different surface minimum unit, according to the cover class in the land.
- Urban Fabric and Water bodies: 1 ha.
- Agricultural land: 2 ha.
- Forest and Natural Areas: 2 ha.
- Wetlands, Beaches, greenhouses, riverside vegetation: 0,5 ha.
• Cartographic scale: 1:25.000. According to final geometric accuracy: quadratic mean error (X,Y) ≤ 5 m. Screen resolution: ≈ 90 pixels/inch
• Harmonized Object Oriented Data Model (OODM), UML notation.
• Update cycle: 2-3 years, according to the Very High Resolution image coverages provided by the Spanish National Land Monitoring Program

2.3.2 SIOSE Data model: using object oriented concepts

As explained before, the hierarchical nomenclature system, used in CLC and other LC databases, has shown its incapacity to store in an efficient way the diversity of LC occurrences in the European territory. Nowadays, conceptual data models, designed using object-oriented methodology, and its conversion in relational or object-oriented databases, are best and common practices to build Information systems. Considering this, Land cover should be described using a harmonized conceptual parametric data model, based on ISO 19109 (Rules for application schemas, ISO TC 211 standards), which classes must be defined and kept in time in order to identify land cover changes within time series. Using this data model, standard European and international nomenclatures (CLC, UNFCCC Land Cover categories according to IPCC guidelines) can be automatically obtained from SIOSE, shaped like thematic queries. OODM SIOSE data model was designed, in collaboration with more than 25 institutions interested in LU/LC, organized in "Thematic Working Groups", similar to Inspire organization. From this OODM, it is possible to derive as many 'views' as necessary. Standard CLC or Murbandy/Moland Nomenclatures would be "predefined views" of the OODM database. If any SIOSE user (local, national or international institutions) has the need to input additional information of a specialized field. E.g.: agricultural, forestry, infrastructures, etc., it is easy to "extend" the Data Model, using one or more of these techniques:

- Subdivide one UML class into multiple classes
- Add more UML classes
- Add values to controlled list text parameters
- Add more parameters to some classes.
- Add "conditions" that parameters should comply.
- etc.

3. CONCLUSION

The main conclusion can be summarized as the need for harmonized capture and production of the environmental geoinformation in horizontal (thematic requirements) and vertical terms (Global, Regional, National and Sub-national requirements). This goal can only be achieved by improving the cooperation and dialogue between data producers, users, and political responsible for the generation of environmental geodata, basically Regional and Global user of environmental data, National Mapping Agencies, Environmental Ministries and Agencies and other related public and private corporations.

Inspire in Europe has developed the appropriate legal framework for this cooperation, giving legal support for the long term actions required for harmonization of environmental spatial data, but further and decisive steps should be done by stakeholders in order to really coordinate and rationalize data production, taken into account that duplication of geospatial data capture should be avoided. There are three pillars for this coordination:

1) Digital Landscape Data Model, able to respond to real National and International user needs and requirements, designed by consensus between stakeholders, using the standardized modelling techniques (ex. UML, Object Oriented design) specified by the ISO TC 211 standards. Once designed, and accepted by users, this Data Model should enable unique data capture, which considers the consolidated common requirements for the main environmental uses and applications.

In SIOSE project, one of the main goals was precisely the Data Model designed by consensus by National and Sub-National thematic expert, which translated their common requirements into landscape classes and parameters. Later, it was easy to coordinate production flows in time using this previously agreed data framework.

2) Use of cost-efficient normalized land monitoring data, mainly imagery (satellite images and ortophotography), as an objective reference data for extracting geospatial information. The development of new techniques for acquiring environmental information, such as LIDAR data, and the extensive use of digital photogrammetric cameras for producing orthoimagery, with improved spectral information (addition of near infrared band in digital aerial photos) enables cost-efficient production of land information with additional biophysical parameters, such as crown density, state of vegetation, and land humidity.

As an example, Spanish PNOT and SIOSE have shown how is possible to coordinate land monitoring in a cost-efficient way using orthoimagery, through a technical dialog and cost-sharing between stakeholders.

3) Improvement in production flows, adding flexibility in funding (ex. Public + Private Partnership), joining existing capabilities in countries, regional institutions (ex Europe) and global (GEO and GEOS) in a double bottom-up and top-down approach, coordinating National and Sub-national environmental inventories production (bottom up) where data can be captured with high geometric and thematic resolution, and adding regional monitoring capacities for improved and cheaper production (ex GMES' Sentinel image coverages). This will harmonize technical and social requirements with real social and budgetary constrains, always taking into account the need for 'free and open' environmental geodata.