

INTEGRATION OF REMOTELY SENSED DATA INTO GEOSPATIAL REFERENCE INFORMATION DATABASES. UN-GGIM NATIONAL APPROACH

A. Arozarena^a, G. Villa^a, N. Valcárcel^a, B. Pérez^a

^a Spanish National Geographic Institute, Land Observation Unit, 3 General Ibáñez de Ibero, 28003 Madrid, Spain - (aarozena, gmvilla, nvalcarcel, bpmartin)@fomento.es

SpS2 - ICSU, Geo Unions, JB GIS, UN-GGIM

KEY WORDS: Remote sensing, UN-GGIM, geospatial reference information, inter-disciplinary, cooperation

ABSTRACT:

Remote sensing satellites, together with aerial and terrestrial platforms (mobile and fixed), produce nowadays huge amounts of data coming from a wide variety of sensors. These datasets serve as main data sources for the extraction of Geospatial Reference Information (GRI), constituting the “skeleton” of any Spatial Data Infrastructure (SDI).

Since very different situations can be found around the world in terms of geographic information production and management, the generation of global GRI datasets seems extremely challenging. Remotely sensed data, due to its wide availability nowadays, is able to provide fundamental sources for any production or management system present in different countries. After several automatic and semiautomatic processes including ancillary data, the extracted geospatial information is ready to become part of the GRI databases.

In order to optimize these data flows for the production of high quality geospatial information and to promote its use to address global challenges several initiatives at national, continental and global levels have been put in place, such as European INSPIRE initiative and Copernicus Programme, and global initiatives such as the Group on Earth Observation/Global Earth Observation System of Systems (GEO/GEOSS) and United Nations Global Geospatial Information Management (UN-GGIM). These workflows are established mainly by public organizations, with the adequate institutional arrangements at national, regional or global levels. Other initiatives, such as Volunteered Geographic Information (VGI), on the other hand may contribute to maintain the GRI databases updated.

Remotely sensed data hence becomes one of the main pillars underpinning the establishment of a global SDI, as those datasets will be used by public agencies or institutions as well as by volunteers to extract the required spatial information that in turn will feed the GRI databases.

This paper intends to provide an example of how institutional arrangements and cooperative production systems can be set up at any territorial level in order to exploit remotely sensed data in the most intensive manner, taking advantage of all its potential.

1. GEOSPATIAL REFERENCE INFORMATION

Geospatial Reference Information (GRI) constitutes the “skeleton” of any Spatial Data Infrastructure (SDI). Its main characteristics are (FGDC, 2005, Rase et al., 2002):

- Provides an unambiguous location for a user’s data (i.e. it is georeferenced).
- Enables merging (aggregation/fusion) of data from various sources.
- Provides a geographic framework or context to allow others to better understand the spatial information that is being presented.
- It is subject to a regular data maintenance regime, with a defined life cycle.
- It is provided by an authoritative source with a mandate (responsibility) for its maintenance and availability, usually a National Mapping and Cadastral Agency (NMCA).
- It is generated at the maximum resolution and accuracy that current technological processes allow.

GRI is hence an official dataset that is reliable, accurate to the level of user requirements, sustainable and homogenous, valid

not only at national level but at regional and global levels as well. It should be generated using a combination of bottom-up (from national to global level) and top-down approaches (from global and continental to national level). Collaborative maintenance of the databases has to be also considered, including the participation of Volunteered Geographic Information (VGI).

The contents of GRI are yet to be precisely defined, being this task under current development. However, some geographic themes can already be identified as GRI: geographical names, administrative units, transportation networks, hydrography, elevations (DEMs), orthoimagery, settlements and land use/land cover information.

Among the constituent elements of GRI, there is a subset defined as “core GRI” (cGRI) which, according to the regional body of the United Nations initiative on Global Geospatial Information Management (UN-GGIM) in Europe, is defined as “*authoritative data from UN Member States that satisfy minimal needs of cross-border, European and Global level*”. This means sustainable and reliable data mastered by countries, upon which thematic data owned by users can be based, and with temporal and historical dimension. It is meant to address supranational requirements with homogeneous specifications and content at pan-European level,

but also capable of becoming the previously mentioned “skeleton” at local and national level. The cGRI must be as objective as possible, avoiding unnecessary subjectiveness. It shall be obtained through mainly automatic methods, within the current possibilities allowed by technological developments.

The flow of geographic information (GI) production and maintenance should start with the generation of cGRI by Public Administrations of the Member States (MS). GRI is obtained by adding other reference datasets to cGRI, whose generation is performed by the private sector and volunteers (e.g. OpenStreetMap). Additional datasets are added to GRI by the public and private sector, generating the commonly known as geographic information (GI). This GI has to be managed using some kind of geographic information management system (GIM).

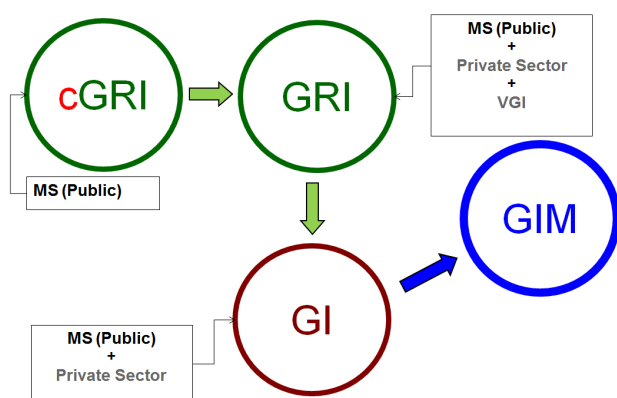


Figure 1: Flow of geographic information as envisaged by UN-GGIM Europe

2. REMOTELY SENSED IMAGERY AND GEOSPATIAL REFERENCE INFORMATION

2.1 Platforms and sensors for land observation

Remote sensing satellites, as well as aerial and terrestrial platforms based sensors, produce every day huge amounts of imagery and other data from a wide variety of sensors. This imagery is not only to be seen as the main source for the extraction of GRI, since they are in their essence GRI. This means:

- Provides an unambiguous location for a user’s data (i.e. it is georeferenced) with a high spatial definition.
- It is relatively easy to update in a reasonable period and at a reasonable cost.
- It is capable of being merged with other geospatial information.

The main sources of remote sensing products are explained below:

- Digital cameras for aerial photogrammetry capture stereoscopic images allowing orthophoto generation at very high resolution with pixel sizes between 20 and 50 cms covering whole countries and 10 cms and even 5 cms in areas of special interest, such as urban or coastal areas. Due to the amount of applications requiring these datasets (topographic and cadastral mapping, land cover/land use, agricultural, forestal and hydrological studies, among many others)

most countries have governmental programs that produce these types of coverages with a defined periodicity, which usually varies from 2 to 4 years.

- Very high resolution optical satellites (between 30 cm and 1 m) from a variety of commercial satellites allow to update the aerial orthoimage coverages in the intermediate years, especially in urban areas and areas of potential change
- Airborne LiDAR sensors capture georeferenced point clouds (X, Y, Z values together with the intensity of the returned pulse and other ancillary data) of objects on the ground. From this point clouds, it is possible to generate Digital Elevation and Surface Models. Point clouds are also very useful in a wide variety of applications, such as flood risk modeling, detection of unregistered buildings, vegetation studies, range of telecommunication antennas, etc. As in aerial photogrammetry, many countries already have implemented programs to periodically capture aerial LiDAR coverages of their territory. Continuous acquisition of areas of interest allows the comparison between two point clouds of the same area in different times, obtaining changes in the territory in a semi-automatic way, as changes in elevation can be considered in most cases as changes in geographical sense. In many countries an increasing will of implementing national programs of periodic acquisition of airborne LiDAR data has been detected, due to the favorable cost-benefit relationship.
- Mobile-mapping systems obtain multi-view georeferenced images together with very high-density point clouds. These datasets can be used in transportation networks (roads, railways, etc.) and their surrounding areas to extract geometry and attributes such as axes, limits, vertical and horizontal signals, slopes, track width, state of road bed, etc. More and more countries are implementing plans for covering complete transportation networks with these systems.

2.2 Cost of acquisition and cost-benefit relationship

Medium-resolution satellite images (considering medium resolution the interval ranging from 10 to 100 meters) has begun to be distributed free of cost in the last years by several spatial agencies (ESA, NASA, INPE). Therefore there is a growing interest in many countries on making the best use of this data, and hence initiatives such as Collaborative Ground Segments (CGS) for the Sentinel missions are being adopted by Member States in the EU.

On the other hand, acquisition of coverages by other types of sensors imply higher costs but once they are captured, these coverages can be distributed to other social agents, namely companies, associations, NGOs and citizens. By openly distributing the data to these agents, the number of applications increases exponentially, together with the return as benefits for the society.

2.3 Data policy

Open data policies are prevailing nowadays. In particular, this type of policy greatly benefit VGI initiatives (OpenStreetMap and others) since they are based on the altruistic work of non-professional people. However, this type of information must be supported by data captured by the above-mentioned sensors, acting as the basis underpinning the rest of additional information. As a consequence, these types of data need to be provided by public authorities.

2.4 Ground truth and Volunteered Geographic Information

One of the most difficult needs to be satisfied when extracting information by semi-automatic processing using the previously

mentioned data is the definition the ground truth areas needed to “train” the algorithms and validate the results.

This is a need where synergies with VGI can be very fruitful, since volunteers can be an important source of ground truth, focusing in those areas where they live and have detailed information.

2.5 Massive processing and Big Data

In the last years, major developments in all sensor types allow to apply algorithms and processes to the data much more sophisticated than the traditional ones. The drawbacks of these developments are the download time as well as the needs for data storage, which grow exponentially as the number of available sensors increase. Many Terabytes need to be processed using much more demanding algorithms in terms of computing power.

This issue, referred to as “Big Data”, occurs in all science fields, in which has become mandatory to rapidly process big amounts of data. The most basic principle of Big Data is to invert the traditional paradigm of downloading the raw data to a local computer and process it in the same location. The new paradigm is to upload the required software or application to the location where the data is stored and make all the processing in that location, downloading only the resulting dataset.

Several technologies have been developed to allow these processes: Virtualization allows processing in “virtual machines” (VMs), “hypervisors” that emulate a physical computer, in which an operating system can be installed and configured, together with all needed applications and libraries. VMs allow “overcommitment” by creating several VMs in one real machine (RM), with the sum of resources higher than that of the RM. This enables to make use of the “simultaneity coefficient” (in a given moment, many VMs will be not performing any processing) and hence, they will not consume any resource. These VMs can be cloned as many times as needed, with all software installed and configured, and made easily available to users on “marketplaces”.

Parallelization allows to use many machines (even thousands of them) to perform in a collaborative way a certain task, achieving the result in a shorter time. Cloud computing is based on the “collaboration” among several RMs forming a “cluster”. This allows to offer almost unlimitedly scalable IT services (storage, processing) without the user being owner or administrator of the hardware constituting such “cloud”. These technologies solve the issues of implementation, and allow creating “private clouds” in the facilities of a given organization, or “public clouds” in machines of a cloud service provider that will only charge users for the storage and processing power actually used. In most cases this is much less expensive, than acquiring, administrating and maintaining an owned processing center. “Mixed clouds” can also be created that will only use external services in processing or storage peaks.

3. NATIONAL INSTITUTIONAL ARRANGEMENTS AND THEIR IMPACT ON THE GENERATION OF GRI

3.1 Co-funding and institutional arrangements

In many cases, the costs of data capture cannot be covered by a single organization, given the high costs per square kilometer and the need of frequent updates. Besides, the investment would be difficult to justify for just one of the many applications of the data.

However, given the high number of potential applications (especially when data is offered to all agents in a free and open way) if an inventory of the needs and requirements of all interested parties is done, followed by consolidation in a co-funded common project satisfying most of them, with proper technical specifications, the cost-benefit ratio results greatly favorable. The investment is then justified, resulting in great benefits for the society.

Adequate institutional arrangements have to be put in place to match the requirements and organize the co-funding.

3.2 United Nations Committee of Experts on Global Geospatial Information Management

United Nations initiative on Global Geospatial Information Management (UN-GGIM, <http://ggim.un.org>) is a global initiative, launched in 2011, aiming at playing a leading role in setting the agenda for the development of global geospatial information and to promote its use to address key global challenges.

The Committee of Experts in UN-GGIM is mandated to provide a platform for the development of effective strategies on how to build and strengthen national capacity on geospatial information, as well as disseminating best practices and experiences of national, regional and international bodies on geospatial information related to legal instruments, management models and technical standards.

3.3 National Institutional Arrangements Working Group

The Spanish National Geographic Institute (IGN) is involved in the activities of this initiative as coordinator of the UN-GGIM Working Group on National Institutional Arrangements (NIA WG, http://ggim.un.org/UN_GGIM.wg4.html), which intends to identify best practices, sets of institutional models and legal frameworks for national geospatial information management and interoperability between different systems and institutions responsible for its management, while ensuring uniformity and standardization. The institutional models should provide governments with options on best practices to create national geospatial entities.

To accomplish this primary objective, NIA WG is analysing geospatial information management organizations all around the world from different political, technical, economic and administrative points of view. Its works, focused on building an index on best practices based on a set of meaningful indicators, will follow a three-step process, from the characterization of the current situation to the classification of the different organizational systems according to the proposed index.

In order to develop this Work Programme three Task Groups (TG) were established in order to tackle specific scopes, namely:

- TG1: Production systems analysis.
- TG2: GI funding structures, dissemination systems and data policy models.
- TG3: Structure of GI management organizations and the role of VGI.

3.4 The Spanish case

Spain is a country with a de-centralized government system, where three types of government coexist: a general state administration, 17 autonomous regions and 2 autonomous cities, and local governments. All these agents participate in the data capture and

generation of GRI.

Geographic data is captured through national plans for land observation, thanks to the use of photogrammetry and remote sensing. Since 2004 this activities are coordinated through the National Plan for Land Observation (PNOT), in which the central administration (through several ministries) and all the autonomous regions participate. It is thus a cooperative program that pools technical, logistical and economic efforts through adequate mechanisms of inter-administrative coordination led by IGN and the National Center for Geographic Information (CNIG). They are sustainable projects, based on user's requirements and with a long-term shared financing scheme. They allow maintaining cooperation strategies in all areas, enabling the increase of added value products and innovation in geographic information.

This decentralized and cooperative production approach between different administrations agrees with the spirit of the European INSPIRE Directive.

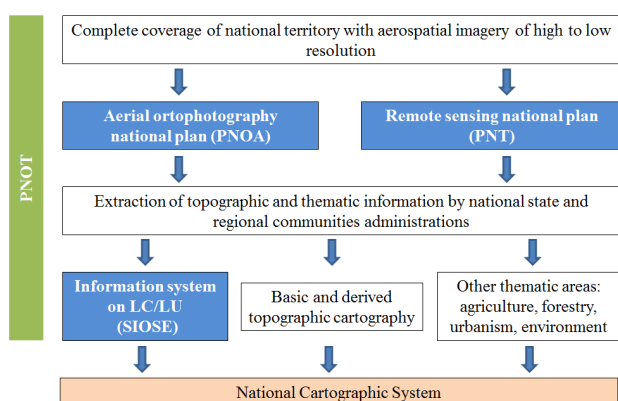


Figure 2: Spanish National Plan for Land Observation (PNOT)

In the first phase, the objective of PNOT is to obtain coverages with satellite and aerial sensors for a wide variety of applications, with economically optimized technical specifications and updating periods, and to develop and promote applications in which these data are going to be used. These coverages are organized in different levels of spatial detail and time frequencies, which are materialized into specific plans:

- The National Plan for Aerial Orthophotography (PNOA, <http://pnoa.ign.es>) is aimed at obtaining digital aerial orthophotos of 25 or 50 cm resolution and high precision Digital Elevation Models of the whole Spanish territory, with a three years updating period are also produced. PNOA is a collaborative project co-funded between the central administration and the autonomous regions. Orthophotos with 10 cm resolution of urban areas and the coast (PNOA 10 cm) and LiDAR coverages of the whole national territory complement PNOA. Download of data and non-commercial as well as commercial uses of PNOA data are free and with no cost.
- The National Remote Sensing Plan (PNT, <http://pnt.ign.es>) provides regular coverage of the Spanish territory with current and historical satellite imagery. This imagery is acquired with multi-user licenses for all Spanish public institutions (stakeholders include several ministries and all the autonomous regions) and processed once with geometric and radiometric processing chains agreed by experts of the

Spanish scientific community. PNT imagery is structured in four levels:

- Very high resolution: images from 30 cm to 1 m of urban areas and the coast.
- High resolution: Images from 1 to 10 m of spatial resolution of the complete territory each year.
- Medium resolution: Images from 10 to 100 m each 16 days.
- Low resolution: Multispectral images from 100 to 1000 m of spatial resolution, each 1 or 2 days.

To reduce the time between the image capture and the availability of the information, PNT has designed a processing system, archiving and online dissemination using adequate servers and networks. There are some important aspects such as designing a storage infrastructure suitable to the volume of information, an appropriate workflow, distribution control and an efficient dissemination.

The second phase ingests the data produced in PNOA and PNT to multiple applications by the various participating organizations, producing basic and thematic geographic information in order to carry out service and development activities. The generation of a detailed Digital Elevation Model is the necessary first step, followed by the extraction of the hydrographic network using automatic methods. In order to generate the GRI, it is mandatory to look for consistency between geometry and topology at different resolution levels. Coherence is necessary between all basic layers of the GRI (transportation network, urban settlements, land cover/land use, etc.).

Imagery from PNOA and PNT are also used in the generation of the Spanish Information System on Land Cover and Land Use (SIOSE, <http://siose.es>). SIOSE is a long running project which integrates different data from regional and national administrations. It has become the Spanish and European reference as regards geographic information. SIOSE has produced since 2005 a multidisciplinary spatial data infrastructure, periodically updated, in a decentralized and cooperative production model by Spanish national and regional administrations. This projects has been 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 as well as their environmental parameters. This complex and strong description of each land cover occurrence allows obtaining environmental and land key indicator from SIOSE data. SIOSE is also one of the main precursor of the EAGLE concept, an object-oriented conceptual solution and set of tools to support a European information capacity for land monitoring built on existing or future national data sources according to INSPIRE principles (Arnold et al, 2013).

4. CONSLUSSIONS

In the last years, the establishment and development of the global initiative on geospatial information management UN-GGIM, and the promotion of the use of this type of information has revealed of great importance in order to tackle the most important global challenges.

NMCA agencies should provide a common system for land observation and GRI production, to address the needs of public administrations and users in general. In order to satisfy these requirements, it is necessary to produce, at least, the cGRI in a

continuous and sustainable manner. Hence, GRI production and its associated processes (dissemination, updating, etc.) should become a priority for most public organizations. In addition, the schedule and the total/annual costs should be evaluated and assumed in advanced by all stakeholders.

Another contribution to the production and updating of GRI would be the implementation of collaborative platforms able to allow distributed and collaborative editing of GRI databases by all involved public organizations.

Finally, GRI becoming the same and unique “skeleton” of GI would avoid duplicities of efforts and expenses thus entailing great economic savings (estimated in more than 60%). GRI should be reliable information, guaranteed by the Member States. Intellectual property of GRI corresponds to the states, while the optimal data policy would be free of charge download and use.

REFERENCES

Arnold, S., Kosztra, B., Banko, G., Smith, G., Hazeu, G., Bock, M., Valcarcel N. 2013. The EAGLE concept: a vision of a future European Land Monitoring Framework. EARSel Symposium proceedings 2013, “Towards Horizon 2020”.

FGDC, 2005. Framework Introduction and Guide Handbook. Digital version, <http://www.fgdc.gov/framework/handbook>.

Rase, D., Björnsson, A., Probert, M., Haupt, M. (Ed). 2002. INSPIRE Reference Data and Metadata.

Peces J.J., Villa G., Arozarena A., Tejeiro J.A., Domenech E., Plaza N. 2010. Spanish National Remote Sensing Program, a way to achieve massive use of remote sensing data. 3th RAQRS, Valencia, Spain.

Valcarcel N., Arozarena A., Villa G. 2011. Breaking down barriers between cartography, geospatial information and environmental monitoring data: towards a new production model. 25th International Cartographic Conference, Paris.

Valcarcel N., Villa G., Arozarena A., Garcia-Asensio L., Caballero M.E., Porcuna A., Domenech E., Peces J.J. 2008. SIOSE: a successful test bench towards harmonization and integration of land cover / use information as environmental reference data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B8. Beijing 2008.

Villa G. NATO science for peace and security series C: Environmental security. Chapter in “Geospatial visual analytics”. 2009. Cooperative decentralization: a new way to build added value chain with shared multi-resolution satellite and aerial imagery and geoinformation.