DATA HARMONISATION AND OPTIMISATION FOR DEVELOPMENT OF MULTI-SCALE VECTOR DATABASES

M. Bernard [a,*], T.Rousselin[b], N.Saporiti[b], M.Chikhi [c]

[a] Spot Image, 5 rue des Satellites, BP14359, 31030 Toulouse, France - marc.bernard@spotimage.fr
[b] Thierry Rousselin & Nicolas Saporiti , Geo212, Paris, France - www.geo212.com
[c] Beijing Spot Image, Beijing, China - myriam.chikhi@spotimage.com.cn

KEY WORDS: Multi-source, multi-resolution, vector database, data Kits, imagery, requirements and constraints

ABSTRACT:

The increasing number of Earth observation sensors on the market is making the choice of optimized data sources to serve operational needs more complex than ever before. Applying a geographic perspective before selecting data sources and procuring services that offer comprehensive data source kits are two effective responses to this new trend.

The paper presents a methodology newly developed and implemented in the field of building multi-scale vector databases on a national or regional size. One of the main problems when a vector database is to be built is to identify, select, and qualify the adequate input source data, knowing the classes of vector / objects to be captured, and the correspondent technical requirements (eg. minimum size, geometric accuracy, maximum "age" of data,…). The paper shows how a rigorous methodology based upon "traditional" geographical considerations, associated with a focused exploration of the catalogues of image providers, greatly help to select an efficient and as-cheap-as-possible multi-resolution portfolio of input data aimed at minimizing the risks of data capture as well as the QC tasks.

The associated presentation will detail the methodology for the selection of pyramidal multi resolution image layers, and their qualification for vector production, vs the specific scales required. Finally, the authors would like to share the lessons drawn after several years about the technical specifications / requirements listed of the vectors themselves, for new bases as well as for updating existing ones, and bring these few pragmatic elements to the scientific community in charge of developing innovative algorithms for the new generation of tools for vector identification and capture.

1. NEW TRENDS IN MAPPING PROJECTS

1.1 Constraints on Map Production

Operational mapping projects (both military and civilian) have to reconcile contradictory constraints. Areas to be covered are large and located in varied geographic zones where existing maps are most often inadequate. While map production is still mostly scheduled in yearly or multi-year cycles, projects sometimes have to ratchet up quickly to meet short-cycle requirements in line with the evolving economical, social or geopolitical context. The required quality of geospatial products and services dictates the quality and uniformity of data sources selected. And because areas to be mapped are sometimes difficult to access (for security, or time-frame, or economical motives), a major part of the brunt of production and verification is performed remotely. As a result, it is hard to assure high-quality mapping, and the judgment of users working with geospatial products on the field—sometimes months or years after source acquisition and data production—can be quite harsh.

1.2 Mapping in Military Context in the 1990s

All of these factors drove military geographers' early interest in making broader use of Earth observation space assets. However, throughout the 1980s and 1990s this interest was damped by the technical and operational limitations of available solutions. In particular, low sensor resolution and lack of agility precluded building up a coherent spectrum of data sources adapted to all kinds of products and environments. Consequently, military geospatial products derived from the SPOT 1 to 4, ERS-1 and 2 and RADARSAT-1 satellites faced big constraints: the need to plan acquisitions over areas of interest well in advance; turnaround times of months or years, even in zones where tasking was considered easy; limited intrinsic geometric product accuracy and limited ability to extract features from imagery. In addition to these constraints, the need to crank up the entire chain of image suppliers and data producers when required to respond at short notice had a significant impact on project costs.

The stand-out satellite-based products delivered to the French defence establishment during the 1990s were SPOT HRV imagemaps covering entire countries or operations zones, and ERS/JERS imagemaps of French Guiana at a scale of 1:200 000. From 1999, combined acquisitions from SPOT 1, 2 and 4 afforded the ability to generate high-quality digital elevation models (DEMs) and orthoimages for cruise missile guidance. For the so-called "Bison" project, Spot Image set up a system for pre-qualifying SPOT HRV stereopairs to guarantee performance requirements. This successful operation served as a test bed —for the Ministry of Defence as well as for Spot Image, the French Mapping Agency IGN and Istar— for future SPOT 5 data products. It also enabled them in September 2001 to fully meet an operational requirement for 300,000 km² of orthoimages of Afghanistan in three weeks, the first real large-coverage geospatial project of its kind in France's military history to rely on satellite imagery.

1.3 Toward a Service-based Model for Mapping Projects

The arrival of a new generation of satellites —Landsat 7 ETM+, Terra, SPOT 5 (with HRS and HRG sensors), IKONOS, QuickBird and Envisat— made it possible to optimize the technical aspects, costs and turnaround times of the geospatial products defined in the 1990s. It also spawned new products and services.

The biggest shift was toward a service-based model. Today, the French defence establishment no longer buys imagery alone, but a global capability to meet the specifications of complex products like DEMs, land-cover data and multiscale vector databases through internal or outsourced production lines. This guaranty is provided by Standalone, Multi-Source Data Kits, through which suppliers undertake to provide a coherent workflow able to generate a geospatial product or family of products (eg DTED level 2 DEM Products or MGCP vector data base). This system includes :

- A geographical study of the area,
- The Data Kits,
- An exhaustive Qualification of the data,
- A database of relevant links for specific topics (oil & gas data bases, environmental info, urban info,…),
- An image database of characteristic objects associated with recommendations regarding the vector extraction,
- Extended guarantee to face eventual production problems.

These components will be described in the following sections.

This new approach was initiated by the French DNG3D defence programme, starting in 2002 along with the launch of Reference3D©. This product contains a DEM basemap plus a highly reliable orthoimage that serve as a foundation for most subsequent value-added products. Reference3D© exploits the outstanding coverage capability of SPOT 5's HRS instrument. Since 2005, the Topobase Defense operation (within the frame of the Multinational Geospatial Co-production Programme), ensures the industrial-scale production of a complete range of raster and vector geospatial products using imagery from 12 civil and military satellites.

2. STANDALONE SOURCE PACKAGES

2.1 Data Kits for Multi-scale Mapping

Each Kit is a predefined package that addresses specific scale needs and geographical conditions such as climate, relief, population density,… The content of each Kit, designed to allow standalone production over the corresponding area, is as follows :

- image coverage(s) for vector extraction,
- data for geometric registration,
- data to fill gaps (clouds, technical imperfections,…),
- data for quality control of the end product.

This methodology is well adapted to multi-scale geospatial products providing increasingly fine details in areas of high interest. What is more, customers are offered a fixed price per sq.km per Kit, whatever the actual content of the delivery, that makes budget planning much easier .

For military projects, the Kit methodology also enables more effective task sharing between civil and military sensors by fully leveraging their respective performance capabilities.

2.2 Pyramidal Coverage

Kits are classified into 3 categories, on a coverage/resolution/scale/cost compromise basis. As an average after several years and half a dozen large mapping projects, we can state that the usual Kit blend delivered to map a region (ie several hundreds of thousands of km²) is a pyramidal coverage containing typically :

- 100% coverage with a Level 1 Kit (eg Reference3D + Landsat7)…
- … plus 35% to 55% coverage with a Level 2 Kit (eg SPOT 5m or 2.5m Pansharpened) …
- … plus 5% to 10% coverage with a Level 3 Kit (eg Ikonos or QuickBird).

Pyramidal multi-coverage of the middle-density and high-density zones is necessary to ensure an homogenous interpretation of images and thus a good consistency of the final vector products across images, administrative boundaries and…. human operators in charge of extracting the vectors.

Moreover, the very limited amount of Level 3 Kits is key for budget reduction and allows to update more frequently the most demanding areas (indeed, this multiscale Kits approach also helps to better focus updating procedures).

2.3 Including New Sensors within the Kits

Until now, about ten Kits has been designed for 1:250,000 to 1:10,000 multi-scale mapping. But the methodology is also flexible and could evolve to accommodate new sensors as they enter operational service, complementing the existing suite of data sources.

For example, the advent of optical sensors offering a daily revisit capability like FORMOSAT-2 makes it possible to meet tight delivery constraints in zones where difficult weather conditions make it hard to image for SPOT 5. The wide-swath, high-resolution acquisition capability of KOMPSAT-2 will also lighten the workload on other civil and military sensors. Future high-resolution radar sensors such as Radarsat-2, TerraSAR-X and Cosmo-Skymed should provide coverage of zones that preclude good optical imaging, and products of more or less equivalent quality.

And "second-string" sensors like IRS or DMC, although offering a lower level of global performance, can improve revisit times and complement key sensors (eg as gap fillers or QC data) like Landsat 7 or SPOT. Leveraging a complementary suite of sensors in this way makes the overall acquisition system more reliable and robust, while fulfilling new product requests at the same fixed price per sq.km.

3. GEOGRAPHY IS BACK

3.1 Understanding the Geography of the Area is Key to Select Appropriate Material for Mapping

Of course, the combination of application needs and the growing spectrum of sources is spawning a diversity of potential solutions that can seem daunting to the layperson.

Within a few years, we have moved from a situation where users could only hope to find at least one usable image of their area of interest —sometimes at the expense of quality— to the current situation in which most of the time users are faced with a choice between multiple images. This impression became even more fuzzy in the last two years with the emergence of numerous "free" images available via virtual globes or on the Internet.

For this reason, it is time to put geographic analysis back into the mapping project design process and into a more value added product oriented quality control of the imagery. Associated with analysis of the physical environment and human activities, geography gives us most of the tools we need to select the best Kits for operational requirements. We still need the engineering expertise that suppliers like Spot Image and others have long provided to optimize one or more sensors to fit requirements. But we must now combine this expertise with in-depth analysis of the area to be mapped, its features and their impact on our choice of data sources. This analysis calls on methodologies allying open sources, domain expertise and quality assessment of existing geospatial products (general and thematic maps, image mosaics, specialist services, etc.) to inform users' decisions and offer guarantees concerning the products that will be derived from the chosen sources.

This step results in a report that proposes a selection of Kits for the area : each proposition of a Kit is justified by facts and reasons dealing with the specific conditions of the area of interest.

This study provides also 2 other outputs :

- A database of relevant links for specific topics (developing trends, oil & gas, tourism, economic strengths or weaknesses) : relevant papers and Web sites, existing geospatial databases, contacts, … This information proves to be very helpful not only to justify decisions in term of initial geospatial production but is also a good starting point for future enhancements or updates of the data. In military wording, we can say that this DataBase is built from complementary Open Source Intelligence (OSINT) results on the area of interest
- An image database of characteristic objects associated with recommendations regarding the vector extraction. Here again, it is a tool which allows decision makers to weight the amount of effort they will put on a certain area (in terms of cost, time or expected performances of the end product) but it will also be used to familiarize production operators with local specific objects and environment.

Report and databases are thoroughly discussed with the user, and purchase orders for images and satellite tasking are placed only after full agreement is reached. This process does not necessarily generate perfect data but at least all parties are fully aware right from the beginning of the risks and of the impacts of the upstream decisions on the end products, cost and delay.

3.2 Qualification of the Data

This service oriented approach and the results of the geographic studies conducted before the selection of source Kits allow to better focus qualification of the data. Image providers like Spot Image have long established a standard in terms of image quality control, but it is now possible to make a step toward a more value-added product-oriented quality control.
On the top of qualifying "basic" image characteristics and specifications compliance (which of course is mandatory), is the good knowledge of the area makes it possible to focus specific controls towards the vector extraction potential of the images. This is done not in general terms (Do we see the roads on those images) but very precisely (do we see the track in this particular area where we know it will be particularly difficult for the imagery because of local conditions).

Here again, the goal is not to provide perfect source Kits (perfect data never exist) but to establish and present a clear and straightforward relationship from source selection to end product processing.

The result of the Qualification process is a report which tries to predict the practical impacts of the actual content of the Kits on, mainly, the quality of the final products, and the delay. This Qualification process thus comforts the whole workflow to the eyes of all the players, who finally benefit from it.

4. CONCLUSION

The French defence establishment adopted this new approach more than 2 years ago through its involvement in the Multinational Geospatial Co-production Programme (MGCP), but it applies equally well to all kinds of users—urban development agencies in Africa, oil and mining operations, funding organizations—looking to secure the best return on their investment in space-based remote sensing. It is amusing to note that, as technologies in this area are reaching maturity and we are applying the lessons of a 20-year heritage in satellite-based military mapping, we have come around to recognizing a basic truth: good mapping requires (a little) geographic knowledge.