Data management for the Advanced General Survey and Non-Technical Survey Tools

Didier Peeters, Sabine Vanhuysse, Eléonore Wolff
Free University of Brussels
IGEAT

1.   Abstract

The Advanced General Survey (AGS) and Non-Technical Survey (NTS) tools require processing sophisticated and heavy geographical data. A system for managing spatial relational databases is being developed to optimise the storage, exchange and analysis of these data. The nature and general objectives of TIRAMISU (i.e. a toolbox for mine-action users around the world) leads to keep the general cost for users as low as possible, and hence to provide a system that is as open as possible, making it compatible and upgradeable. Therefore, we have chosen to develop the tool using an efficient open-source Spatial Relational Database Management System (Spatial RDBMS) called PostGIS. This system will operate in the background and be hidden for the regular users, but will be accessible to the power users willing to enhance the tools.

2.   Introduction

The mine action activities involve a wide range of technologies of very different nature. Among them some deal with electronic data to support the practical field activities or to detect, localise and prepare to field work. Until now, different actors have developed software to achieve these tasks but often independently from the rest of the community. A well-known effort was made by the GICHD with their tool IMSMA devoted to support Mine Action Centres information management anywhere in the world. Recently studies, like the SADA study [1] funded by the ESA that ended in April 2012, pointed the need for a common data storage system for the demining actors.

In the Geographical Information Systems (GIS) particularly, there is a strong interest in sharing data between different tools to accomplish very specific tasks, but until now the GIS software used in the demining world are based on proprietary technologies depending on private companies and are rather unable to easily collaborate. The tools that are developed each have their own data storage system that is closed to any third-party enhancement.

The TIRAMISU effort consists in providing the demining community with a set of tools, "*… to bring a toolbox that will represent a step forward in mine action by being the basis for a unifying, comprehensive and modular integrated solution to the clearing of large areas from explosive hazards*"[1].

A "*modular integrated solution*" means first that the tools are complementary to each other and to the existing ones, and second that each tool has to be integrated with the others. We consider that data sharing is the key to achieve this goal, and for sharing data the only valuable solution is to organise the data into a common structure, which means having a database that satisfies all the features requests of the different actors.

3.   The GIS database requirements

Inside TIRAMISU the remote sensing and GIS tools are:

1. T-REX, geospatial, contextual and expert information collection and management,
2. T-IMAGE, acquisition of airborne and satellite remote sensing imagery,
3. T-MAP, features of interest mapping from remote sensing imagery,
4. T-SHA, spatial analysis for supporting SHA re-delimitation,
5. T-PRIORITY spatial analysis and multi-criteria methods for supporting high-level mine action management.

---

[1]TIRAMISU Description of Work (October 20, 2011)

These tools handle vector and raster data. Some raster data come from very-high resolution aerial or satellite images that are far too big to be processed by regular computers and will have to be entrusted to third party. They are therefore not considered here.

Beside these tools, the users in the mine-action community already use desktop GIS software, on their own operating systems and the technical solution that will be developed must take this into account.

4.   Respect of the standards

One of the most useful mean to avoid the incompatibility between systems is the respect of standard technologies. It was therefore decided during the first project year to favour the use of standards. Standards are usually defined by consortia of influent actors who represent most of the users of a specific technology. Many standards are actually very commonly used like the World Wide Web protocol defined by the W3C consortium, others are more specific. For instance inside TIRAMISU we plan to rely on:

- OGC for Geographic data definition
- SQL for Database language
- The maXML for the regular mine action data (developed together with the GICHD)
- The web for communication protocol

Complying with the standards ensures a better life expectancy for the TIRAMISU achievements and offers an opportunity for future enhancements by other actors of the demining world.

5.   The database system

For all these reasons, the choice was made to rely on the Relational Data Base Management System (RDBMS) PostgreSQL, with its geographical extension PostGIS.
PostgreSQL is a free open-source RDBMS[2] offering an undoubtedly professional steady system used all around the world in any kind of field. PostGIS adds it the ability to handle geographical data, by adding geographical data types (like points, lines, polygons and their derivatives), and hundreds of functions to process these data.

It is compatible with most operating systems (notably Windows, Linux, OS X) and can be accessed by a wide range of GIS, among which ArcGIS, GeoServer, GRASS GIS, Map Info, gvSIG, QGIS, MapServer, and many others, and of course complies with the OGC standard [2].

The initial use of the database is to support the TIRAMISU tools but it is also possible to read the stored data with the GIS systems mentioned above to produce other output.

PostgreSQL + PostGIS only require a regular PC for normal use but of course, the hardware requirements are more dependent on the data that are being processed, like for any other systems. A big aerial image, for instance, would still require a huge amount of memory once imported into PostGIS, as it would in another software. PostGIS is one of the most efficient spatial data systems but it does not work miracles.

6.   Use of the database on site

The sophistication of the system will have no negative impact for the final users. All the operations will be handled by the tools developed in TIRAMISU, but the database itself will not be visible. This is similar to the use of IMSMA where the regular users are actually not aware of the data storage, organisation or retrieval, they just use IMSMA's front-end.

With our tool, instead of opening files the users will select data in a structured list searchable according to key words or the data classification, and reversely for the storing. So, concretely the TIRAMISU tools will provide the front-end of the database. The "power users" may nevertheless go beyond this and use a dedicated database front-end interface to accomplish some specific tasks like developing specific tools, by using for instance the PostGIS functions.

---

[2] More exactly a "ORDBMS", Object-oriented RDBMS

PostgreSQL is a multi-user system that allows to define different user types, to control the accesses, and fine-tune the access rights for each type or individual user.

Even if Postgres/PostGIS offers numerous "state-of-the-art" functionalities, the first reason for choosing it is the security provided to the GIS operations, in terms of integrity for the data, control of the processes. It also offers tasks automation and possibilities to add new tools or functions developed elsewhere.

During the last years the usual GIS have progressively evolved from file storage systems (like the shapefile) toward RDBMS solutions because of their greater robustness, their ability to handle efficiently almost limitless amount of data and their general faster performances. MS Access, which is a low end DBMS, for instance, has limited capacities in terms of database size.

A professional RDBMS such as PostgreSQL ensures also the integrity of the data in case of incident like a power failure, by reversing an operation that wouldn't have gone until its normal end. Other systems without this feature would generate data that have lost some internal consistency, meaning that an erroneous data could remain and have unexpected consequences some day.

The tools will be able to import or export their data in regular files for exchange purpose with systems different from TIRAMISU. PostGIS and PostgreSQL provide also tools for such exchanges, PostGIS can directly import shapefiles, Arc/Info, kml, geojson, and others (any OGR supported format actually, but not only).

Finally, with the SQL language and/or the PostGIS functions it is easy for GIS educated users to set up their own additional functions to execute tasks that would be needed on a regular basis.

To conclude this chapter we would like to quote an interesting reading document [3] available on internet that summarizes the advantages of the GIS RDBMS this way:

- Convenience - easy access
- Reduce redundancy - cross referencing reduces repeated data entries
- Shareable - cross organization sharing, multiple uses of data
- Flexible analysis options - multiple analysis options, expansion of analysis capability
- Standardization - controlled formats for data entry / manipulation reduce inconsistency and errors.

Below is a graph showing the data-flow between the GIS tools with the database in the middle. This illustrates the role of the database in the data exchanges.
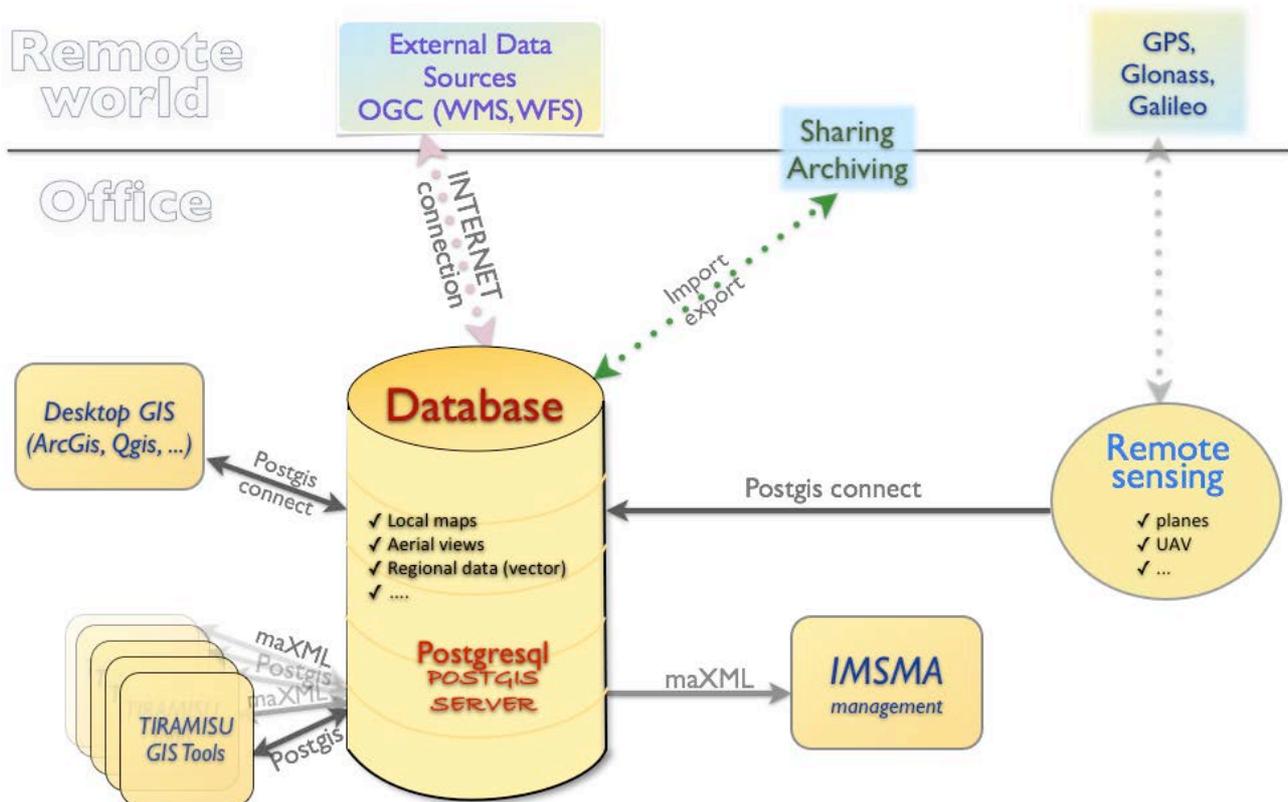
Graph 1 : Schematic view of the data flow with the database unit

7. The database during the project development

The TIRAMISU GIS teams will work simultaneously on the structure of a common database and on their respective tools. This database will be hosted on a server in a location still to be determined and be accessible by secured internet connection. So the database and the tools will be developed in synergy, ensuring a good collaboration between the partners involved. This method will act as a rehearsal for the final delivery.

8. Integration with other work packages inside TIRAMISU

We strongly believe that a centralised database system is an asset for all the software tools and that the idea should be extended to the whole TIRAMISU toolbox, but that is a question to be addressed on another level. Anyway, beside the work packages that address specifically GIS, there are other work packages that might take advantage of the database structure, like those relying on the result of the GIS operations. Unless the development of a database common to all the work packages is undertaken - which would solve this question, the GIS database will be made accessible to the partners who would need it for their own work.

A relational database should be seen like a cabinet in which it would be possible to add drawers or boxes and of which size would be dependent on its content. So it is always possible to add extensions (meaning other stuff) without any major consequences for what is already in it.

General organisation of the data

The "relational" nature of such a database has consequences on the organisation of the data inside it. Relational databases theory has pointed out rules that are real advantages when respected, and have very positive consequences on security, speed and efficiency. This is no place to explain these constraints but let us say that one major characteristic is that the data should be organised according to their nature and not according to who creates them or needs them. The data are stored in objects represented as tables that can be linked together when needed. This avoids redundancy, and hence ensures homogeneity and spares disk space. For example, considering a suspected hazardous area (SHA) identified by some GIS process should not be mixed with the administrative information related to the region in which this SHA is located. Instead there should be a table with the regional administrative information and a table with the SHA and an

identification code for the region, so that when needed a synthetic report can be produced with the SHA description including the name of the region and other information. It is even possible to have this synthetic information presented in a "false" table, called a "view", resulting from a stored SQL code that is executed automatically each time new information is inserted in one of the involved tables. This view looks like a table, it can be used by automatic processes like a table, but it is not a table. So this is a way to have the best of both worlds, on one side a rigorous data structure controlled by strong rules, and on the other side a more "user friendly" result.

9. Evolution capabilities and post-TIRAMISU perspectives

Our mission in TIRAMISU is to make tools that are useful for mine action, and hence for protecting people. After the project lifetime, mine action will keep going on, maybe endlessly. It is therefore our responsibility that the current effort could be a source of improvements and even more, a basis for further developments.

By using technologies like PostgreSQL or PostGIS, that are open, free and rely on standards, we hope that other people anywhere, will see in it an opportunity to develop their own tools, and share them with the rest of the demining community.

As mentioned above, this database system allows writing new tools, as SQL scripts, that can be shared with all the future users of the TIRAMISU GIS database. There is a huge Postgres/PostGIS community in the world very active on internet and sharing the expertise freely. Similarly there are other technologies, like for example "R"[3] that can be connected to Postgres/PostGIS to make analytical operations (statistics, reporting, …) also freely and with the same collaboration possibilities.

We see also a possibility to open our system to other existing complementary systems, like IMSMA, which could take advantage of the data stored in our database.

References

[1] Kruijff M. et al., *Space Assets for Demining Assistance*, Acta Astronautica, Volume 83, February–March 2013, Pages 239-259.

[2] Steiniger S., Hunter A.J.S., *The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption*, Computers, Environment and Urban Systems, In Press, Corrected Proof, Available online 10 November 2012

[3] http://www.cfr.msstate.edu/students/forestrypages/fd/fo4313/topic15.pdf

The PostGIS official documentation : http://postgis.net/docs/manual-2.0/

The PostgreSQL official documentation : http://www.postgresql.org/docs/

---

[3] http://cran.r-project.org