Revista Brasileira de Cartografia (2017), Nº 69/8, Edição Especial "Geovisualização, mídias sociais e participação cidadã: apoio à representação, análise e gestão da paisagem": 1442-1455.
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

SDI AS A CORPORATE TOOL AND SHARING INSTRUMENT: DATAGEO'S ROLE IN SÃO PAULO'S ENVIRONMENTAL SYSTEM

IDE COMO INSTRUMENTO CORPORATIVO E DE COMPARTILHAMENTO: O PAPEL DO DATAGEO NO SISTEMA AMBIENTAL PAULISTA

Christian Rezende Freitas[1], Charles Rezende Freitas[1], Arlete Tieko Ohata[2] & Gláucio Almeida Rocha[3]

[1]Universidade Federal de Minas Gerais – UFMG
Escola de Arquitetura / Laboratório de Geoprocessamento
Rua Paraíba 697, sala 410a, Savassi. Belo Horizonte, MG, cep 30130-140
CHRISTIANRF@GMAIL.COM, CHARLESRF@GMAIL.COM

[2]Secretaria de Meio Ambiente do Estado de São Paulo
Departamento de Informações Ambientais
Av. Professor Frederico Hermann Jr., 345 - São Paulo, SP, cep 05459-900
AOHATA@SP.GOV.BR

[3]Secretaria de Meio Ambiente do Estado da Bahia
Consultant for the Environment Develop Program of IDB
3ª Avenida 390, CAB – Salvador, BA, cep 41.745-005
glaucio.rocha@gmail.com

Recebido em 10 de Dezembro, 2016/ Aceito em 16 de Junho, 2017
Received on December 10, 2016/ Accepted on June 16, 2017

RESUMO

O artigo tem como objetivo relatar a construção do DataGEO, uma infraestrutura de dados espaciais ambientais do Estado de São Paulo, discutindo as definições que nortearam sua concepção e implementação no âmbito do Sistema Ambiental Paulista. O texto aborda as relações do projeto com a INDE (Infraestrutura Nacional de Dados Espaciais), influenciadas pelas especificidades de uma IDE temática, ambiental, e os processos observados no funcionamento da SMA-SP. O DataGEO amplia o horizonte de uso das IDE's não só como plataforma de disseminação de dados espaciais para o público externo, mas também como proposta de solução de ambiente de GIS Corporativo de grandes organizações. Além de apresentar os processos de implementação e os desafios tecnológicos, o artigo visa discutir essa nova forma de acesso à informação georreferenciada baseada em WebGIS, na qual a interoperabilidade e o compartilhamento de dados favorecem o suporte ao planejamento. O projeto DataGEO demonstra a importância de se propor uma plataforma na qual a busca pelo entendimento das diversas demandas à informação, em uma condição acordada de visualização, pode resultar na ampliação do acesso, uso, compreensão e compartilhamento de dados para suporte a decisões coletivas.

Palavras-chave: IDE, Interoperabilidade, WebGIS, Sistema de Suporte ao Planejamento.

ABSTRACT

This paper aims to present a description of the DataGEO project, São Paulo's environmental SDI (Spatial Data Infrastructure), discussing the guidelines used to develop the design and implementation in São Paulo's environmental systems. The text includes the issues around the NSDI's (National Spatial Data Infrastructure) guidelines and the choices made throughout the project, influenced by the specifics of a thematic environmental SDI and the processes observed in the functioning of SMA-SP (Environmental Secretary, State of São Paulo). Thus, DataGEO expands the SDI's use presenting a platform that may be used not only as a solution for the sharing of spatial data with the public, but also as a proposed solution to attend corporate GIS within large organizations. The paper presents the implementation process and the technological challenges, in addition to discussing new ways of accessing geospatial information based on WebGIS, in which the interoperability and sharing of data helps and supports spatial planning. The authors point to the importance of how a proposition for a platform allowing the integration of information coming from multiple sources through institutional agreements may improve access, usage, sharing and understanding data for group decisions.

Keywords: SDI, Interoperability, WebGIS, Data Support for Planning

1. GEOGRAPHIC INFORMATION IN ENVIRONMENT MANAGEMANT

A good deal of time has passed since the environment has been analyzed focusing on the relation between the natural elements of landscape. From a more holistic approach, environmental studies involve the investigation of the interdependence between physical, biological, anthropological and socio-economic conditions.

Due to the global impacts of environmental changes, countries have developed a series of protocols that have established goals seeking to reduce those impacts. The environmental laws of each country are often a mirror of these protocols. Under Brazilian legislation, specifically, each governmental branch has its own legislation, respecting the responsibilities of each branch.

Therefore, the laws that oversee the topic in question should encourage sustainable development, while attempting to reach a balance between economic, social and environmental aspects and the interests of the government, society and the market.

As Tauk-Tornisiel (1995) points out, these aspects as a whole require different variables be evaluated at the same time under different criteria, for example some of them should be measured in US$, while others in g/m^3. As a result, some authors have considered the importance of environmental studies and territorial planning as mechanisms of identification, control and management as areas of direct conflict on economic growth and environmental quality.

There is a fundamental element needed for this model to work: information. The viability and the guidelines for interventions in the environment are determined by collected and analyzed data. As these data are generally presented in different formats, it becomes harder to analyze as a whole. Therefore, maps become a useful alternative, because they put the elements of data on a level comparison basis, providing geographic appeal for this kind of information, where data is almost always connected to a location in the territory. Thus, Geo Information Systems (GIS) has become one of the main models for storage and use of environmental data today.

When consulting some of the literature on this subject, Xavier-da-Silva (2001) states that the concept of environment as a system does not consider the processes that generate environmental phenomenon, but also brings to light the necessity of the identification of important environmental characteristics, such as geographic position, the territorial extension (that leads to the concept of shape) and spatial and functional relations.

Moura (2003, 2014) also discusses the power of geographic information on environmental studies reporting that the advantage of using spatial modeling is on the promotion of resources that help with the spatial visualization of phenomenon, since the spatial make-up may result in new interpretations of spatial outcomes.

Both Moura and Xavier da Silva highlight the importance of a systematic approach on environmental analysis, with the creation of outlooks and models in digital format, using GIS. These are the tools that allow the manipulation of geographic data, their analyses and the publication of results.

The products of mapping, particulary those generated with the application of remote sensing, such as satellite images and aerial surveys, are the main source of registering and obtaining these changes in space and time, contributing to the use of spatial information in environmental investigations.

Therefore, the increase in the demand

of spatial data with the intention of building environmental models naturally increase. This demand is the interest of entrepreneurs, government agencies, companies specialized in Geotechnologies, environmental studies and society – each one maintaining its own interest in the environmental process.

It is noteworthy when observing the cycle of geographic information from the point of view of the environmental department, as they are both suppliers of official data to the general public and consumers of information to support their supervision, control and planning of activities. The cycle remains on-going within the departments. The same may be said of the internal processes of these departments, in which the information produced by one department is the gateway to projects of other departments.

Recently, a Brazilian law named "Information Access" (Law 12.527, 2011), has set out legislation stating that government institutions must now publish unrestricted public data, forcing institutions to adopt new mechanisms for publishing and sharing data. Geographic information is included among the categories mentioned in the law.

2. SPATIAL DATA INFRASTRUCTURE

An SDI (Spatial Data Infrastructure) can be understood as a platform to organize, search and share spatial data. As Coleman (1998) and others point out, this platform is composed of five main elements: people, data, institutions, technology and technical standards.

By people, we understand all the consumers and producers of data that use the platform. Data is the geographic information itself, whether in vector, raster or document formats. The institutions are the organizations tasked with the creation of adequate policies and the allocation and maintenance of the necessary infrastructure. The technological element is composed of tools such as the metadata catalogue, the map servers, the DBMS (database management system), WebGIS and others that allow for the discovery and access to information for both users and external systems. The standards are the means to guarantee the interoperability, offering users user-friendly, open-file formats, placing special attention on the standards of the OGC (Open Geospatial Consortium) and ISO (International Organization for Standardization).

It is paramount to understand how an SDI works, highlighting how the elements interact (Fig. 1).

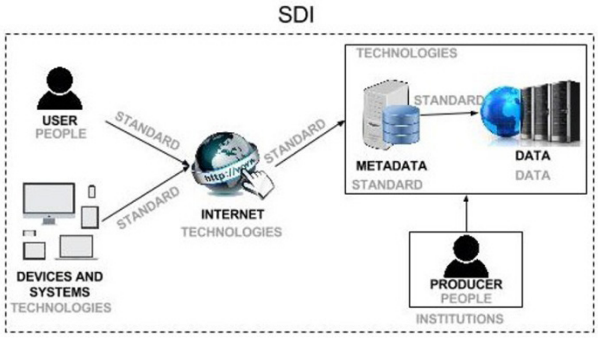

Fig. 1 – Elements of a SDI. Source: Produced by the authors.

The SDI attempts to reduce the time, work and the resources required to obtain data, avoid redundancies in information and update constant format changes, thus avoiding the use of outdated information.

As regards the limitations and the relation within the SDI, Rajabifard (1999) believes they may vary on a global scale, involving institutions with both a global presence and less detailed data, and a local level, involving institutions with less spatial scope, yet more detailed data. According to this theory, the infrastructures relate to one another, as the lower levels contribute to the formation/constitution of the higher-level infrastructure systems. There are different levels and they have specific relations with one another (Fig. 2).

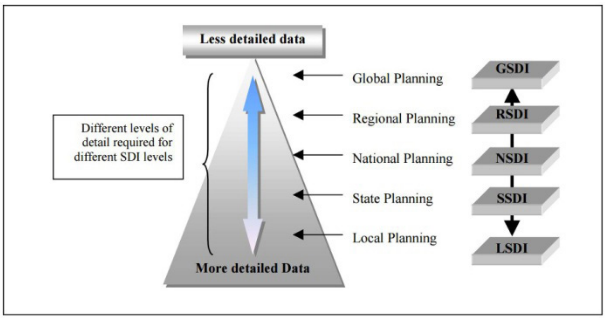

Fig. 2 – Relations between levels of SDI. Source: Rajabifard et al. (1999)

To illustrate this point, INSPIRE, a European Spatial Data infrastructure, can be used as an example. It has been composed of national infrastructure networks that are also composed of other SDI's, both local and regional.

It is common to find authors that include the corporate level of SDI's, connected to private organizations, while others may consider it as a local level resource.

It is also common to find a categorization according to the subject when the infrastructures involve well-determined themes such as geology, environment and other issues. According to Orshoven (2003), SDI's targeting users, that need specific subjects, can be classified as TSDI (Thematic Spatial Data Infrastructure) and may be positioned on any level according its diversity.

From an evolutionary view, SDI's have been created resulting from the need to share geographic information, borne mainly from government agencies. Fueled by the growth of internet in the 90's, information sharing using file transfer protocols such as FTP (File Transfer Protocol) grew exponentially.

In the following decade, we can see a migration of this file structure method to one based on website interface, dedicated to the distribution of geographic data, although mainly focusing on supplying data in file formats linked to commercial Desktop GIS software. As a result, the use of service standards through the internet made it possible to develop SDI's in the format as we know them today.

In a classification system, as proposed by Davis (2006), the infrastructures are separated in first and second generations, depending on the way they interact with the user through the metadata. According to some authors, the first generation makes direct reference to the data, even using metadata, offering file downloading. The second generation, employs the concept of SOA (Service Oriented Architecture) to provide access to data, using OGC (Open Geospatial Consortium) web service standards. This difference offers a variety of opportunities to the construction of infrastructures that use multiple data sources and machine-to-machine connection (Internet of Things).

In Brazil, the initiative to build a NSDI (National Spatial Data Infrastructure) started with legislation on November 28, 2008, Decree 6.666, tasking CONCAR (National Cartography Commission) with the mission to plan out implementation, including activities until 2020.

The Brazilian NSDI (National Spatial Data Infrastructure) has developed a handful of documents that have led to other initiatives to build SDI's within the country, with special attention to BGM (Brazilian Geographic Metadata) and EGDV (Technical Specification for Geographical Digital Vector Data Modeling) for NSDI and e-Ping, the electronic government interoperability standard that specifies the OGC pattern as reference to the federal government's geographic services.

These specifications have been based on international standards employed around the world, bearing a few modifications aiming to promote the interaction between SDI's on different levels and subjects.

3. DATAGEO PROJECT

The organizational structures of SMA-SP, São Paulo's State Environmental Agency, are composed of departments, research institutes and foundations. This agency includes more than 48 offices in the countryside, in charge of research, inspection, control, licensing and environmental planning. They consume and produce large amounts of geospatial information that are used by the agency itself, in addition to the community that deals with this type of information.

The DataGEO project aims to supply SMA-SP with a platform that helps with the territorial management as it organizes and shares geospatial data, not only for institutional use, but also for public access. This platform is an environmental spatial data infrastructure.

The other functions of DataGEO aim to complete and support the TSDI in institutional and technological issues.

The main benefits of a TSDI are:

- Organize and unify the institution's official geographical data.
- Improvements in acquisition of geographical information, avoiding buying the same information twice and rationalize efforts to create new data.
- Cut costs in hiring services; the official data is available, therefore increasing the quality of the information produced.
- Rationalization of internal processes and information flow, due to the standardization in the use of geographic

information.

- Cost-reductions in developing new geographical information systems, since TSDI resources will be dedicated to the resolution of specific problems.
- Freedom of technological choices and decrease in annual costs with licensing and infrastructure maintenance.

We can classify DataGEO as thematic, organized on a state level and second generation. Its creation has required a series of technical and technological definitions from the group responsible.

The technical choices are related to the project's management with the definition of metadata and service standards and the data modeling.

These choices focus mainly on tools such as the geoportal, the metadata catalogue, WebGIS, map servers, the DBMS and other tools.

3.1. Methodological choices

During the creation of the project, the team employed an interactive, incremental approach, compatible with the concept of cycles or sprints, using agile methods. The cycles have created useable products, prioritized with the oversight of the SMA-SP team, adding value immediately at the outset of the project. These so-called quick/fast victories ensured project validation and support, balancing stakeholders expectations and enabling the identification and correction of mistakes as quickly as possible.

Another important issue of note, was the use of the "rolling wave" concept. It enabled initial planning resulting from fewer restrictions and a less detailed scope, as reference material or literature was relatively scarce, thus allowing planning to be improved and streamlined as the project developed.

The choices connected to the metadata, the data and the web services were guided by the standards adopted by the NSDI, however, some difficulties in making the national data characteristics (strongly oriented to systematic cartography) compatible with the needs of SMA-SP and the thematic data, led to a few modifications.

Overall, DataGEO uses the profile of metadata presented in BGM, focused on summarized profiles, adding some elements to allow the direct use of the GIS catalogue according to the tools of choice. The data categories, also connected to the BMG, required a lot of themes to better represent the differences of detail between each map. In addition to the differences in scale, the environmental theme has been well-defined between means and processes. The users of these segments expect to find the data according to categories that represent the biological, anthropological and physical environment and in the processes of licensing, supervision, control and planning. The BGM's data categories were extended to attend the needs of DataGEO.

The data model represented a challenging aspect of making the requirements of DataGEO compatible with the standards as defined by NSDI. The national infrastructure system is directed to systematic cartography, while the environmental theme focuses on geoprocessing, thus requiring the spatialization of laws and thematic maps. According to these considerations, a choice was made not to adopt EDGV as reference to DataGEO. Regarding the service standards, NSDI's recommendations were completely followed.

All of the adjustment related to the national infrastructure were carried out in compliance with international standards as determined by ISO and OGC, to maintain the interoperability of the environmental infrastructure as regards other file formats, tools and systems.

3.2. Technological choices

From the point of view of technological tools, an SDI oriented to services must contain a metadata catalogue, in compliance with ISO-19115 and must offer a catalogue service-web OGC (CSW), a map server that offers data in OGC (WMS, WFS and WCS) standards, and a WebGIS. Other tools such as DBMS may also be part of a SDI.

The choices made had few premises to be considered. The SMA-SP used the SQL Server 2012 from Microsoft as a DBMS; other choices took into consideration the hundreds of existing ArcGIS Desktop licenses in the agency.

Although it is uncommon at the outset of an SDI, the geographic ETL (Extract, Transform and Load) toll was extremely important to

DataGEO. It was used for standardization and established data collecting processes with the producer, loading the SDI's data to the DBMS, automating maintenance and the updating of information.

DataGEO also offers security and administrative devices that were developed using Java technologies, a proxy to control private access and a data source manager used to create different visualization methods.

According to the data path from producer to user DataGEO has the following tools:

- Geographic ETL - Geokettle
- DBMS SQL Server
- Metadata Catalogue – ESRI
- Map Servers – Geoserver, Mapserver and ArcGIS Server
- Data Source Manager – Developed with Java
- Security Proxy – Java
- Map Manager – Developed with Java
- Geoportal – Liferay
- WebGIS - Developed by Coffey
- ArcGIS catalogue plug-in - ESRI

The technological architecture of DataGEO presents the use of all these tools (Fig. 3).

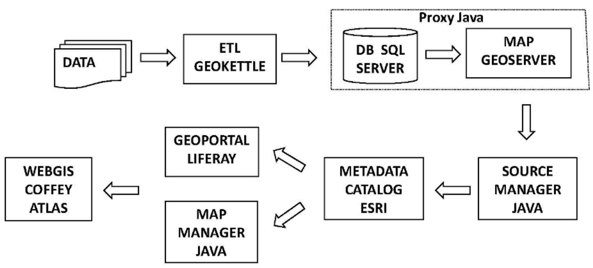

Fig. 3 – Data flow in DataGEO's technological architecture. Source: Produced by the authors.

4. GEOGRAPHIC INFORMATION

One of the main components of an SDI are the group of geographic information that feeds into its technological structure. The challenge in the development of a TSDI (Thematic Spatial Data Infrastructure), such as DataGEO, is in the diversity of themes to be collected and analyzed searching for basic information on the environment and its spatial representation.

The creation of the SDI started identifying the areas that consume and produce information, as concerns the environmental management within the state of São Paulo.

A series of interviews and studies were conducted in order to understand SMA-SP's role in the main business areas, identifying the objects and phenomena that the agency handles, on a level of detail that allows the comprehension of the necessities for the development of an initial concept for DataGEO's data model.

In addition to the concepts and knowledge about the objects and the phenomena related to the environment, it is important that the model also takes the concepts related to the creation of a SDI and the good practices of software engineering into consideration.

As a strategy to build this concept model, work began with the technical SMA team, making arrangements to visit the SMA's business analysts, mapping the path that data goes through, closing circles and identifying the main existing interfaces in the internal institution's relations.

This approach made it possible to identify:

- Main business areas
- Staff members and their respective roles in the business
- Geographical information produced
- The existence of geographical information needed
- Path of information
- Tools and information systems that interact with the geographical information

Meetings were scheduled with representatives of each business area and scheduled by their managers. Information was collected during meetings about business processes, geographic data, softwares and systems used and other additional information.

After compiling the data gathered in the interviews, a diagram was created for each area in which the business process and its future interface with DataGEO can be seen in the figure below (Fig. 4).

In sequence, we attempted to determine the information categories of DataGEO, according

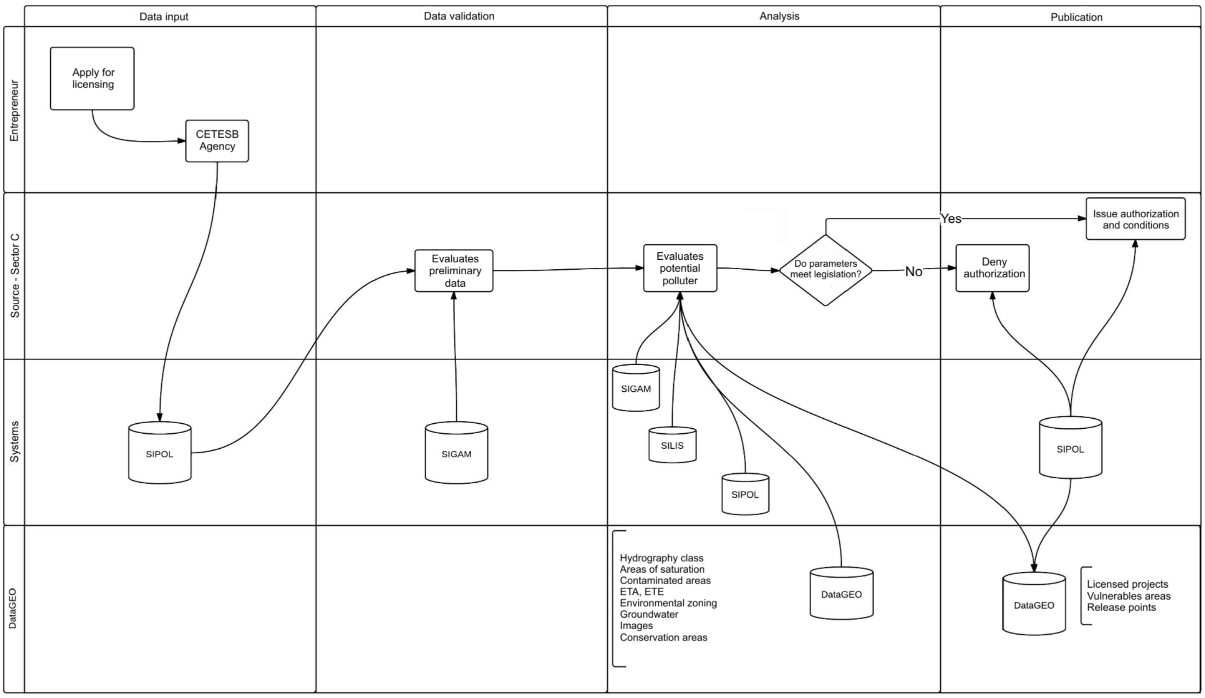

Fig 4 – Example of data flow diagram. Source: Produced by the authors.

to the data gathered and the Brazilian geographic metadata profile documents, BGM, published by CONCAR.

4.1. Information categories

The SMA-SP department oversees a number of activities as regards the environment such as licensing, supervision and management.

As a result, there are a great number of individuals involved with the general processes of licensing, supervision and management and specific processes, such as environmental planning and control of polluted areas that are both creators and consumers of geographic information.

It therefore becomes ever important to be able to identify and have access to the individuals that are in charge of handling this information; therefore, identifying and defining the thematic categories that house this array of information is one the most important attributes of data infrastructure.

It is important to note that the information categories presented below reflect an initial proposal using generic knowledge, and the characteristics could and should be altered accordingly throughout the evolution of both the project and the people responsible for handling it. The evolution of the environmental agency in its task of managing the actions that impact the territory of São Paulo also contributes to change the categories of information. Therefore, these initial categories are dynamic and should always be reviewed by those staff managing the SDI (Fig. 5).

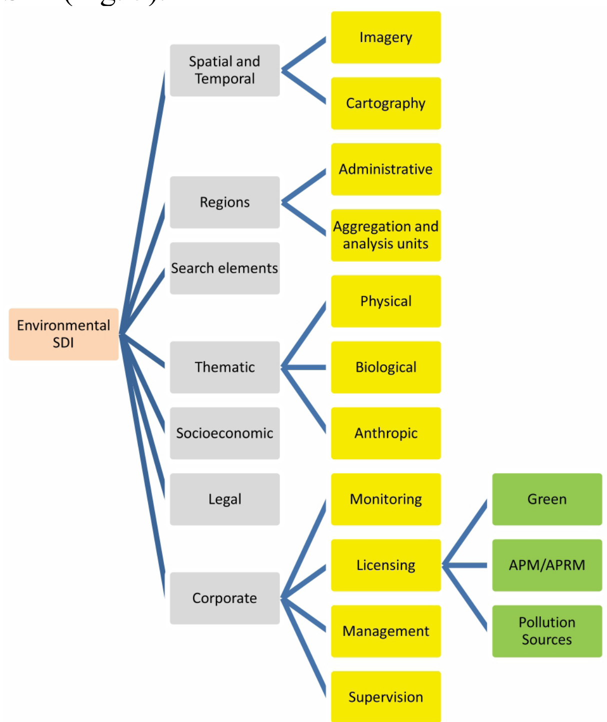

Fig 5 – Categories of information in the conception phase. Source: CPLA (Environmental Planning Department)

It is worth mentioning that the categories of information must be represented in the metadata model making it easier to store and recover information. In order to define the list of categories in DataGEO, the ISO 19115 and BGM profile were taken into consideration.

For the definitions of DataGEO's categories, issues related to the environment and the process of supervision, monitoring, licensing and planning were chosen and classified in a total of 28 categories:

- Water supply
- Restricted areas
- Protected areas
- Energy
- Fauna
- Environmental supervision
- Geology and mineral resources
- Geomorphology
- Hydrology and hydrography
- Imagery and Orthophotos
- Laws and specifications
- Environmental Licensing
- Borders
- Political and administrative issues
- Location
- Basic Mapping
- Environmental monitoring
- Environmental planning
- Surveying
- Natural risks
- Sanitation
- Socio-economic issues
- Soil
- Transportation
- Vegetation
- Zoning and Orders
- Anthropology
- Environmental control

5. HANDLING OF INFORMATION

After identifying the basic groups of spatial information in SMA-SP, the next step was defining how these data were going to be modeled, handled and stored. As a note, the wide variety of information made it difficult to deal with data simply or through single-pattern organization models.

5.1. Data modeling

A data model aims to systematically represent the understanding of objects and events. This representation reflects an abstraction of the real world, which should not depend on the implementation of technologies.

In general, the models represent entities, their attributes and the relationships between them. At present, the UML (Unified Modeling Language) is the main modeling standard used in information system projects.

The representation of geographic entities requires specific items such as an indication of the graphical primitive (point, line, or polygon) representing spatial features. In addition to issues of representation, there are different types of relationships among entities; in other words, spatial relations. These relations are defined from the geometry and geometrical relationship between them: outside, inside, near and others.

Some techniques and models for GIS applications have been proposed, particularly throughout the 90's. They are:

- IFO for geographic applications (Worboys et al., 1990)
- MODUL-R (Bédard, 1996)
- GeoOOA (Kösters, 1997)
- GMOD (Oliveira, 1997)
- GISER (Shekhar, 1997)
- MADS (Parent, 1999)
- GeoFrame (Lisboa and Iochpe, 1999)
- OMT-G (Davis Jr.2000; Borges, Davis Jr. et al. 2001; Borges, Davis et al. 2002; Davis Jr., Borges et al. 2005)

Nearly all of these models use the UML to create specifics in the modelling of geospatial applications. The OMT-G (Object Modeling Technique – Geographic) has been consolidated as the main representation model.

The EGDV (Geographic Vector Data Specification) publication, released in 2007, has been adopted by the Brazilian Institute of Geography and Statistics, IBGE, as a reference for modeling and representation of the NSDI. In this publication, the national committee for structuring the national digital map archive has presented expanded modeling for geospatial data as presented in the OMT-G.

The analysis of the models and standards presented in the publication do not respond to the necessities demanded from environmental spatial information in SMA-SP. The model restricts the use of geographic information for map production, hindering other objectives such as GIS.

The definition of mandatory attributes, forms of representation, different spatial relations from those proposed by the OGC and the excessive number of spatial restrictions create difficulties for specific applications. The DataGEO project, as stated in chapter 3, targets the construction of integrated thematic information for environmental management within the state of São Paulo.

Taking these limitations into consideration, the DataGEO project has adopted OMT-G and OGC spatial relations specifications also used by ISO 19.125:2004.

5.2. Types of data

The database gathered for DataGEO consists of three basic types: geographic, alphanumeric and documents.

For the geographic information the basic formats are:

- Raster files in GeoTIFF format
- Vector files in shapefile format
- Traditional alphanumeric tables
- Geographic tables

The documents were converted to PDF format aiding in the transfer of these files to the internet preventing any deformation in the content.

The DataGEO server was used to organize the data in directories created to facilitate access.

For each item of data incorporated in the DataGEO, a document has been created with the main information about its origins.

The geographic data and the alphanumeric data also present the ETL process required for creation, transformation and loading into the database.

5.2.1. Raster files

GeoTIFF is the chosen file format for raster data in DataGEO. The files are organized in georeferenced format covering the area of interest. In these instances, a grid of scenes must be created to facilitate the identification of specific files. The grid should contain, when possible, a link to directly download the data.

The grids are stored in the database in order to make it easier to create devices allowing queries to identify the situations.

5.2.2. Vector files

In DataGEO, vectors files may be stored in two distinct formats, shapefiles or geographic database (Microsoft SQL Server native format).

The criteria used to define the type of storage are:

- File size – large files are stored in shapefile format to save memory space in the geographic database;
- Importance – files that have little relevance in the context of the data, according to the interviews, were stored in shapefile;
- Topology inconsistency – files that have problems with topology, however their importance is fundamental to the environmental process in the agency were stored in shapefile.

The shapefiles were stored in directories, while the geographic database used Microsoft SQL Server.

ETL Processes

The majority of data and its potential to increase caused by the nature of the subject matter required strategies specifically created for its routine.

Loading of the data into the SDI was a continuous activity throughout the project and it was planned into the workflow of data production and its respective metadata.

The solution chosen was spatial ETL software, GeoKettle version 2.5, based on the suite Pentaho Data Integration (Fig. 6).

This tool is configured to understand different formats and convert data, using rules defined in the planning stage, seeking to comply with official standards, and load them into the geographic database, according to the data modeling.

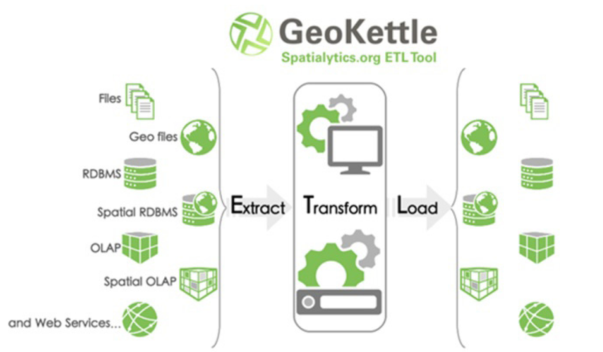

Fig 6 – Diagram of GeoKettle spatial ETL software. Source: http://www.spatialytics.org/projects/geokettle/

The tool choice took into consideration:

- Being compatible with different geospatial data formats and having operations and transformations capable of using topologic concepts for validation.
- Being scalable and allowing access from different locations. Instances should be initiated to help in periods with increase in demand. Capable of working in cloud computing environments or in dedicated servers.
- Allowing improved integration with a service-oriented SDI and the ability to expose transformation activities such as WPS (Web Processing Service) making it possible to register in the metadata catalogue.
- Executing tasks of extraction or creation of automatic geographic metadata in the catalogue.
- Allowing customization, preferably, in Java language, used in the other technological components of DataGEO, unifying the development team.

All the scripts created represent the unique ETL processes. They were tested, approved, documented and stored in a database.

As an example of the ETL process, we can observe the process applied to the creation of aerial security areas around the airfields in São Paulo. In this example the ETL is responsible to extract the points of airfields and create a buffer zone that defines the safety area, loading new data into the database. (Fig. 7).

6. METADATA

Metadata can be understood as information regarding a specific data, a description of how it was produced, by whom, when and other information about its production. When we handle geospatial data, the metadata must contain the details of its origin, bear mapping parameters, such as the reference system, coordinate systems, precision, in addition to describing content.

It is evident that the metadata play an important role in data quality control, not only leading to better results, but also reducing costs by avoiding file duplication and facilitating the exchange of information.

The metadata in SDI is essential, as it is accessed by the user to interact with other parts of the SDI, identifying which data was produced, how it was produced and, more importantly, how it should be used properly. The metadata provides the user the ability to download original data.

The metadata standards are provided in xml format, describing geographic data and its characteristics. There are three metadata schemas used for geospatial data: the Dublin Core, the FGDC and ISO.

6.1. Dublin Core

Developed by DCMI (Dublin Core Metadata Initiative) this standard was published in 1999. It was created initially to verify the content of the information presented in websites such as videos, photos, etc. However, it can be expanded to other kinds of data such as geographic information.

The information to be registered in Dublin Core Metadata was broken down and categorized into fifteen elements:

- Title – data title;
- Creator – name of the individual creating the data;
- Subject – the main theme of the data;
- Abstract – a short description of the content of the data;

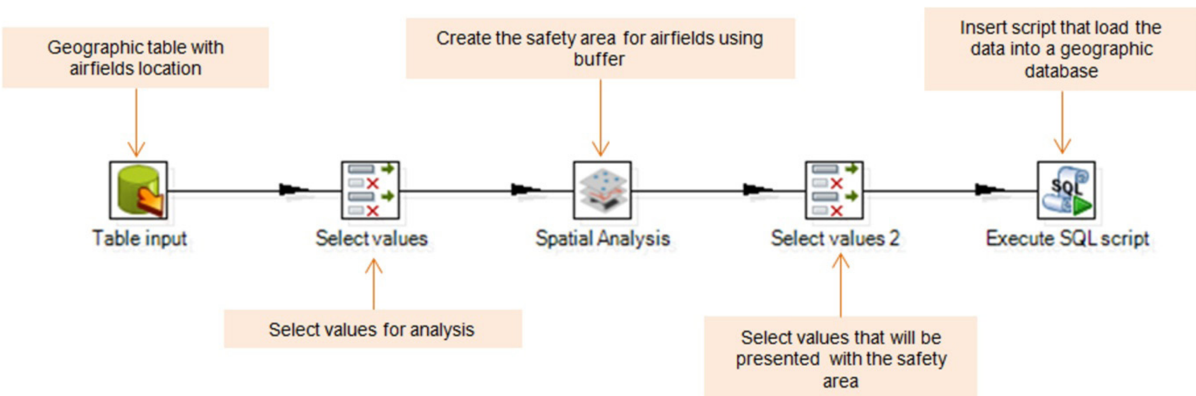

Fig 7 - ETL applied to create aerial security areas. Source: Produced by the authors.

- Editor – individual responsible for the distribution of the data;
- Contributor – entity or people that help in the creation of the data;
- Date – the date associated to an event in the life-cycle of the data;
- Type – the nature of the content of the data;
- Format – physical format of the data;
- Identification – Non-ambiguous identification of the data;
- Source – a reference to other data that may relate to the creation of the present data;
- Language – the language of the content of the data;
- Relation – a reference to other data that may relate to the subject of the present data;
- Extent – the spatial extent of the data;
- Rights – information about the intellectual property rights of data

All elements of the Dublin standard are optional and can be repeated as many times as necessary making this metadata very versatile.

The Dublin Core is not the first standard of metadata developed, even for geographic data, however, it has brought important advances as a result of its proliferation on the web, making it the front-runner of the main models developed since its release.

6.2. FGDC

The FGDC (Federal Geographic Data Committee of United States) is responsible for coordinating all development, use and sharing of geographic data in the North American territories, as well as the American National SDI.

Since its creation, the FGDC has developed a series of policies that have become a reference for the field of geotechnologies, both in the USA and the world.

In 1994, the committee developed a specific standard for geographic metadata, the CSDGM (Content Standard for Digital Geospatial Metadata). This standard contains eleven main sections:

- Metadata – Metadata information;
- Identification Information;
- Data Quality Information;
- Spatial Data Organization Information;
- Spatial Reference Information;
- Entity and Attribute Information;
- Distribution Information;
- Metadata Reference Information;
- Citation Information;
- Time Period Information;
- Contact Information;

Each section contains several entities, and each of those entities has many elements, thus increasing the level of complexity and details in this model.

In 1998, the FGDC overhauled this model and launched the CSDGM v 2.0, in which the main elements remain, and the new system establishes mandatory criteria and detailing with even more subsections of the metadata structure (Fig. 8).

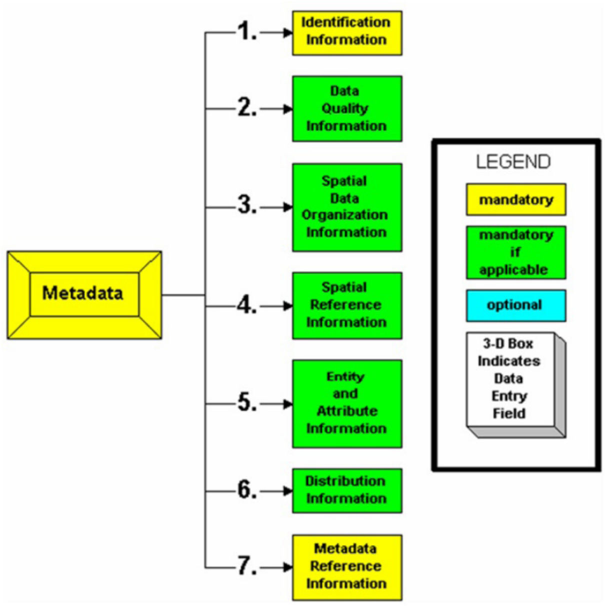

Figure 8 – CSDGM v 2.0 (Source:FGDC)

The FGDC approved the ISO Standard, by the end of 2010, for geographic metadata in the US, as well as in other countries.

6.3. ISO 19115:2003

The ISO proposed a set of compliance standards in the hopes of addressing the quality of geographic data known as ISO:19000. Among the specified standards, the ISO:19115:2003 establishes a standard for geographic metadata.

This standard uses UML to represent sections, entities and elements, creating approximately 400 elements for metadata characterization. Despite the range and complexity, this model has been adopted as a global standard for geospatial data applications.

The main elements are listed below:

- Identification – generic information

about the digital geographic data in which the metadata applies;

- Reference system – description of the spatial reference used by the data;
- Distribution – information about the responsible for the distribution of the data;
- Graphical representation catalogue – information about the rules for graphical representation;
- Metadata restrictions – lists the restrictions for the metadata access and use;
- Application schema - information about the conception schema used by the geographic data;
- Spatial representation – spatial representation of the digital geospatial information;
- Metadata extension – information describing the extension applied to the metadata rules;
- Metadata maintenance – the frequency of metadata updates and its methodology;
- Quality – evaluation of the data quality;

While there are wide range of elements existing in the metadata, most are not mandatory, which is an important factor in making the model more flexible. This enables the data to be shaped in many different situations according to particulars such as in environmental studies.

In Brazil, the NSDI, as well as many others around the world, has become the adopted standard in creation of the BGM (Brazilian Geographic Metadata), with some adaptations defined by the Brazilian Cartography Committee.

6.4. BGM = Brazilian Geographic Metadata

Aiming to establish a metadata standard to be used in NSDI, the Brazilian Cartography Committee evaluated some of the existing standards in the ISO 19115:

- MIG – Portugal's geographic metadata information;
- NEM – Spanish Metadata Center;
- NAP – North American Profile (USA/Canada);
- LAMP – Latin American Metadata Profile;
- IDEP – Basic Metadata profile of Peru.

After the evaluation, the Brazilian Cartography Committee established adaptations from the ISO standard and published the BGM – Brazilian Geographic Metadata, compatible with ISO 19115:2003 in 2009.

BGM considered some information as mandatory to register the metadata and includes specific types to fill in some elements. The summarized profile was:

- Title – mandatory;
- Date – mandatory;
- Author – mandatory;
- Geographic extension – conditional;
- Language – mandatory;
- CDG code – conditional;
- Thematic category – mandatory;
- Pixel size – optional;
- Summary – mandatory;
- Distribution format – mandatory;
- Time and Altimetry extent – optional;
- Spatial representation type – optional;
- Spatial Reference – mandatory;
- Lineage – optional;
- Online access – optional;
- Metadata identifier – optional;
- Metadata standard optional;
- Metadata version – optional;
- Metadata language – conditional;
- CDG code – conditional;
- Responsible for the metadata – mandatory;
- Metadata date – mandatory;
- Status – mandatory.

6.5. DataGEO's metadata

As a TSDI (Thematic Spatial Data Infrastructure), DataGEO goes through specific subjects related to environmental studies, many of them related to geographic planning, requiring the production, transformation and storage of geospatial information. For example, hydrographic data can be set out in a simple thematic map in order to locate rivers, as well as be used as a network structure allowing the use of mathematical models in hydrological studies.

The presence of specialized data is relatively common in a TSDI. This kind of data is used in specific subjects and may have different rules depending on the context in which they are used. These particularities can require expansion or adaptation of the metadata model to attend the global issues of a NSDI.

In DataGEO, the ISO 19115:2003 profile has been adopted, bearing some alterations aiming to improve the data representation and business developed by SMA-SP.

The metadata used in DataGEO have the following elements:

- Title - mandatory;
- Date - mandatory;
- Abstract - mandatory;
- Author - mandatory;
- Role of the responsible individual of metadata - optional;
- Keywords - mandatory;
- Spatial representation type - optional;
- Scale - optional;
- Language – automatic;
- Thematic categories - mandatory;
- Spatial extent – automatic;
- Lineage - optional;
- Reference system – automatic;
- Distribution format - optional;
- URL – automatic;
- Hierarchy - automatic;
- Metadata identifier – optional;
- Metadata date - automatic;
- Metadata name - automatic;
- Metadata version - automatic;
- Individual responsible for the metadata - mandatory;
- E-mail – mandatory.

7. INSTITUTIONAL ARRANGEMENTS

Contrary to a large part of SDI implementation initiatives, in which the arrangements, resolutions and regulations generally precede technical specifications, the DataGEO arrangements occurred after the execution of the initial stage of the project. The current reality of departments and the data produced influenced decision making, and, consequently, the definition of a set of rules established by the technicians before the intervention of higher authorities. DataGEO is currently operational, but agreements are still in the process of regulation.

During the execution of the project, both technical and managerial levels established agreements between the departments in the hopes of cooperation, respecting the autonomy of each department responsible.

After making adjustments and agreeing upon protocols for sharing and the stability for distribution of data through DataGEO, the final arrangements were drawn up and carried out.

This inverse methodology proved to be very effective, as it offsets the authorization for exchange of data according to the current specifics and maturity of each department within DataGEO, thus avoiding misunderstandings among higher authorities.

8. RESULTS AND CONCLUSIONS

DataGEO's design was able to predict the natural evolution of the TSDI (Thematic Spatial Data Infrastructure) from a sharing platform to a component of all GIS and situation rooms in the SMA-SP.

The evolution of the use of geographic information in many departments of SMA-SP, offered a variety of ways to employ the resources of the TSDI to support information systems that already exist or may be built in the future.

The development of information systems by SMA-SP, with the assistance of DataGEO, saves both time and increases the quality of the data created by these applications, since the official data is provided by the SDI. The exchange of information between systems is facilitated through the standardization of geographic information.

Geographic information provided by an SDI is used to compose maps with integrated data created by different systems and departments, visualized by DataGEO.

The incorporation of functionalities provided by an environmental and spatial data infrastructure can be useful in creating information systems that must be aligned with the standards adopted by SMA-SP.

DataGEO represents a tool to validate the geographic information and help support the laws, regarding the procedures carried out by SMAP-SP in environmental management throughout the state of São Paulo.

Evolution in the geographic services provided by DataGEO may guarantee the application of these rules in a standard format.

An important role of the official information of SMA-SP published in DataGEO has been created in the form of legacy systems. The development of an automatic publication interface may be necessary to achieve the goals of SMA-SP: transparency, integration and the sharing information. Those interfaces should be built according to the DataGEO's directives.

The legacy systems, or those that may be built, could enhance the technological advances implemented in the platform that supports the SDI to solve problems in storing, performance or security in future operations.

Sharing information using geographic services, the use of the ETL platform, data standardization, the correct modeling of data, the use of DBMS, the improvement of map representation as a result the use of cache and the use of Cloud Computing specifications are some examples of project recommendations that can be used together with the infrastructure and the methodology developed for DataGEO.

The open technological architecture of DataGEO's open technological architecture and design offer the construction of mobile applications based on the services provided by the TSDI.

Projects that may be implemented in this field will demand updating and adaptation of the services provided by DataGEO, taking into consideration the requirements demanded by these applications.

The gains brought by DataGEO as a corporate GIS tool has surpassed expectations, as it provided a set of tools used by SMA-SP to execute the daily tasks of environmental management with integrated and standardized data and has become an alternative to register the spatial and time changes in the data created by the agency.

ACKNOWLEDGMENTS

To all that directly or indirectly contributed to the construction of DataGEO:

Dr. Clodoveu Davis Júnior – DCC-UFMG

Dr. José Alberto Quintanilha – Poli – USP

Hebert Ramos – Consultant,

Aline Salim, Ana Maria Neves, Cynthia Lina Yassumoto, Edgard Joseph Kiriyama, Igor André Cubateli Redivo, Juliana Amorim da Costa Matsuzaki, Kenzo Matsuzaki, Renata Sayuri Kawashima.

REFERENCES

CAMPAGNA M. **Public Administration Gi-Based Web-Sites For Spatial Planning: a Comparative Analysis**, Italy, 2005.

COLEMAN, D. J. e MCLAUGHLIN, J. 1998. Defining global geospatial data infrastructure (GGDI): components, stakeholders and interfaces. **Geomatics Journal**. 2, 1998, Vol. 52.

CRAGLIA M. and CAMPAGNA M. **Advanced Regional Spatial Data Infrastructures in Europe**. European Commission; Joint Research Centre; Institute for Environment and Sustainability, 2009.

DAVIS, C. A. e ALVES, L. L. Infraestrutura de dados espaciais: potencial para uso local. **Revista Informática Pública**. 1. Belo Horizonte : s.n., 2006, Vol. 8.

KINGSTON, R. Public participation in local policy decision-making: the role of web-based mapping. **The Cartographic Journal,** 2007, 44(2), 138-144.

LEI Nº12.527 de 2011. Casa Civil. [Online] 18 de novembro de 2011. [Citado em: 01 de 11 de 2016.] http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.

MOURA, A. C. M. **Geoprocessamento na gestão e planejamento urbano.** Rio de Janeiro, Interciência, 2014, 286 p. (1a. edição 2003).

OLIVEIRA, P. A., DAVIS JUNIOR, C. A. , OLIVEIRA, P. F. A . Proposição de infra-estrutura de dados espaciais (SDI) local, baseada em arquitetura orientada por serviços. In: X Brazilian Syposium on GeoInformatics, 2008, Rio de Janeiro (RJ). **Proceedings of the X Brazilian Symposium on GeoInformatics.** Porto Alegre (RS): SBC - Sociedade Brasileira de Computação, 2008.

ORSHOVEN, J. V. S**patial Data Infrastructures in Europe: State of Play Spring 2003**. Leuven , INSPIRE, 2003.

PEUQUET, D., MARBLE, D. **Introductory readings in Geographic Information Systems**. London, Taylor & Francis, 1990, 371 p.

RAJABIFARD, A., CHAN, T. O. e WILLIAMSON, I. P. The Nature of Regional Spatial. **27th Annual Conference of AURISA**. 1999.

TAUK-TORNISIELO, S. M., GOBBI, N., FOWLER, H. G. **Análise ambiental: uma visão multidisciplinar.** São Paulo, Unesp, 1995, 381 p.

XAVIER-DA-SILVA, J. G**eoprocessamento para análise ambiental**. Rio de Janeiro, Jorge Xavier da Silva, 2001, 228 p.