MINIMAL DATA BASE ORGANIZATION FOR MUNICIPAL ADMINISTRATION – SUPPORT TO GEOPROCESSING IMPLEMENTATION FOR CITY COUNCILS

Ana Clara Mourão Moura
Charles Resende Freitas
Renata Hungari de Oliveira
Universidade Federal de Minas Gerais - UFMG/IGC
Av. Antônio Carlos 6627 – 31270-901 – Belo Horizonte - MG, Brasil
anaclara@ufmg.br, charlesrf@ufmg.br, renata.hungari@gmail.com

Abstract

This document discusses methodological guidelines for supporting the implantation of Geographic Information Systems in cities, as part of the project *Proex-Mec-Cidades* 2007, of the Ministry of Cities. In the end of 2007, the Ministry of Cities, in collaboration with the federal universities, initiated the *Poext-Mec-Cidades*, a project with the goal of facilitating municipal technicians in the use of geoprocessing. It is a program that incorporates projects of university involvement, with an emphasis on aiding public and social agents in the development of support actions for the state and municipal public sectors, aiming for the institutional development and implementation of information systems that allow the elaborating of plans and projects of urban development, as states the legislature of the City Statute and the National Politic of Urban Development (NPDU), having as objectives to support projects of university collaboration that would contribute to strengthening the NPUD, committed to training technical staffs from public administration as well as social agents, for planned and interactive management. UFMG trained technicians from meso-regions of the metropolitan areas of Belo Horizonte and of Jequitinhonha. As an integral part of the project, it was a goal, not only of preparing the technicians for their first uses of the geoprocessing tools, but also to support them in their firsts steps in constructing minimal cartographic databases for the municipals. The procedure, structured as methodological guidelines, was nicknamed "*how to begin*", and this is this article's object of study. The foci of the present article are the related projects aimed at the training of municipal technicians in the use of the geoprocessing, through courses, the designing of Internet sites for self-learning, elaboration of classes and video-lessons and the building of a methodological guideline. Methodological guidelines were constructed in the format "How to start" with the objective of giving support to municipal technicians in organization principles and the treatment and construction of cartographic and alphanumeric databases consisting of databases of geographic information systems. All the guidelines are presented in simple language and consider, exclusively, the use of free software and data that can be found on the internet. The guide presents information on how to identify the demands on municipal management, how to search for data – maps and tables, how to organize the data, how to convert and adjust the data and organize the products in a Geographic Information System. It's a way to support a beginning of the use of GIS in common practice and to create a culture of organizing

spatial data and spatial analysis. All this information so far discussed must be stored in an orderly way, so as not to lose ourselves with so much information. Create a folder to store all raw data and another to store all alterations, by-products and derivatives from the original files. It is common to produce selections, thematic maps or editions that may serve for other projects. In all file folders, create a text file describing the origin of each file, including how they were produced, scale and origin as well as any other information available about them. We have demonstrated, through the present text, that using only free software and data available on the internet, it is possible to initiate the process of structuring GIS for municipal managing.

1. Introduction

This paper presents a practical experience in the applying of GIS facilities on municipal data control, providing a new way of supporting urban planning. The GIS proposals were first used a long time ago, in the 60's, but we can not say that they are been widely used on urban planning all over the country. In Brazil we still have a great task to spread this technology among an expressive number of city administrations. Before accepting this task we developed scientific studies to build the methodology that we proposed to the project. Among others, our references are Moura (2003), Moura et all (2007), Moura and Borges (2009).

In the end of 2007, the Ministry of Cities, in collaboration with the federal universities, initiated the *Poext-Mec-Cidades*, a project with the goal of facilitating municipal technicians in the use of geoprocessing. It is a program that incorporates projects of university involvement, with an emphasis on aiding public and social agents in the development of support actions for the state and municipal public sectors, aiming for the institutional development and implementation of information systems that allow the elaborating of plans and projects of urban development, as states the legislature of the City Statute and the National Politic of Urban Development (NPDU), having as objectives to support projects of university collaboration that would contribute to strengthening the NPUD, committed to training technical staffs from public administration as well as social agents, for planned and interactive management.

UFMG trained technicians from meso-regions of the metropolitan areas of Belo Horizonte and of Jequitinhonha. As an integral part of the project, it was a goal, not only of preparing the technicians for their first uses of the geoprocessing tools, but also to support them in their firsts steps in constructing minimal cartographic databases for the municipals. The procedure, structured as methodological guidelines, was nicknamed "*how to begin*", and this is this article's object of study.

2. Work's methodology

The methodology was organized in a way to construct a guideline applicable to the actions of implementing geoprocessing in the Mayors Offices, starting from a position without cartographic data bases. As directed by the guideline, the technicians must initiate the work by surveying municipal demands, and by listing existing maps and tables, as well as characterizing their conditions.

Once the existing archive is defined, it must be treated in such a way as to be able to be incorporated into Geographic Information Systems. The technicians are told how to proceed with analogical maps, how to adjust the existing digital maps, how to maximize tables of alphanumeric data and how to incorporate data in the form of tables of points. Special emphases is given to the adjustments of systems of projection and coordinates, as well as to the conversion of formats, a very frequent challenge among those who need to compose their databases from sources of very different origins and nature.

Following the sequence of the guidelines, the technicians are orientated in how to obtain new data, in the case the existing archive is not adequate for their goals. The guideline was organized for practicing the use of GPS in generating road maps, through axils and their respective intersections, with consideration of applying topology knowledge in the collection of the information; the search for obtaining and optimizing the data that is available today online and the elaboration of topographic databases. The technicians are then trained in the organization, structuring and storing the geographic databases, containing the information produced in previous phases.

3. Results and discussion

The guideline's structure is presented below, it is composed of the surveying of demands, treatment of data, acquisition of data and its organizing in the composition of the data bases of Geographic Information Systems.Initially, to discover the demands of each Secretariat, the following steps were programmed:

- To identify the functions and obligations of each Secretariat.
- To identify routine projects and operations executed by each of them.
- To identify obstacles to the execution of tasks.
- In meetings with representatives from the Secretariat and community, listen to the items highlighted as necessities to be addressed and / or problems to be faced.
- Evaluate if each element identified as a demand may be represented spatially (able to be represented in a map) or organized in a database.

3.1. Gathering existing data

To consolidate the database, it is important to verify the existence of information, studies or maps with the Secretariat and other relevant institutions. Along with maps, we also searched for any information in the form of tables or that could be organized as tables.

3.1.1. Maps

The maps may be in an analogical format (paper) of digital (computer files). For paper maps, we must observe some details, such as: dates of production, the source of the information, scale, existence of a coordinate system, *datum* references and, finally, if the map is original or a copy, as well as what kind of copy.

Cartographic databases in digital format may be divided in two groups, image format maps (*raster*) and vector format maps (produced by CAD or SIG software). For both we must observe the following: existence of metadata – which means data about the data, containing name of maker, source used, the scale of the map and the origin of the information.

The image files generally present the extensions *.tif, *.jpg, *.bmp, *.gif, etc. The files known as bitmaps (.tif and .bmp) keep the best image quality of the original product, however, they occupy larger memory, while the more compacted formats (.jpg, .gif), make some colour generalizations and occupy less memory. The *.tif and *.jpg files can be loaded with most of the geoprocessing programs.

The vector files in general present the extensions *.dxf, *.dwg, .dgn, *.shp, *.tab, *.mif and others, programs of the CAD type, are always stored in single files. However, the files SIG's are never alone, but accompanied by several files like *.shx, *.dbf, *.map, *.mid, etc, since SIG files have data of a mutable nature, structured as points, lines and / or polygons, as well as alphanumeric data. Files *.dxf are recognized by most geoprocessing programs and, therefore, are considered formats for exporting and importing (drawing exchange format).

3.1.2. Tables

When we display information in a table, we imagine a structure organized with lines and columns, where information regarding an object, phenomenon, etc can be described. These are the analogical or digital databases that present data as - or can be organized in a - tabulated manner. The digital files for tables may be presented in *.mdb, *.dbf, *.xls and *.csv formats, but the text formats such as *.doc and *.txt are also common. The analogue files can follow a classic lines and columns structure, but may also be integrated as texts. With this type of information, it is important to identify the fields (column titles), the amount of registers (number of lines), the format of each type of field information (text, number, data, currency, etc.) and, in the case of text files, to verify the possibilities of conversion in to tables. Observe the structure of Figure 1.

FIELD 01	FIELD 02	FIELD 03
REGISTER 01	XXXXXXXX	XXXXXXXX
REGISTER 02	XXXXXXX	XXXXXXX

Figure 1. Basic structure of a table.

3.2. Data treatment

Once the appropriate cartographic and alphanumeric data collection is selected, it is important to submit them to treatment processes, since the data is rarely already adjusted to the conditions demanded for putting together the Geographic Information System database.

3.2.1 Map treatment

a. Analogical maps

Paper maps must be filed digitally as soon as is possible. In order to do this, they may be digitalized on large format scanners, which can be found in specialized print shops and stationers, producing a single image of the whole. Alternatively, small scanners can be used to produce several smaller images which may be edited together to reform a whole. In both instances you must define the resolution of the exit image. We recommend avoiding resolutions lower than to 150 dpi (dots per inch), and superior to 600 dpi. We used, in most cases, 300 dpi, since it is the same as working with 0.2mm on the map's scale, which corresponds to visual accuracy. As for image format, we recommend *TIF* for greater image quality. In the case of using smaller scanners for digitalization, one image must necessarily contain overlapping portions of neighbouring images. In cases where it is not possible to identify a scale on the map, we must obtain it from field measurements and using equation 1:

$$Scale = \frac{D_{real}}{d_{map}} \quad (1)$$

By identifying on the map a few points that are accessible in the field, with the help of a GPS, we may determine these coordinates and, consequently, the distance between them (the value of the $\mathbf{D_{real}}$ in the formula) and, with a regular ruler, we obtain this distance on the map (the value of $\mathbf{d_{map}}$ in the formula).

In cases where there is no a coordinate system in the map, we may, with the help of a GPS, determine one in the field, by gathering coordinates easily identified in the map. In this case, the map is given the coordinate system the field worker utilized in the field. The more points that are used, the better the system's definition. In order to finalize the entering of digital maps in our system, we must perform the register procedure of the georeferencing of these images, using the coordinates from the map, when available, points acquired in field with GPS, or better still an already available georeferenced database for the same place.

b. Digital maps

It is a common assumption that files in digital formats are ready for using. In fact, what we observe is that most of these files present inconsistencies and require close analyses. The preparation of digital vector files must begin with topological verification, which means, identifying the correct construction of the objects in SIG representations (point, line, polyline, polygon). We name this process topological cleaning.

Another common situation is the receiving of non-georeferenced vector files. In this case, the same careful observation of image files is applicable, since it is necessary to acquire points, either in the field with GPS or from an already existing database, and then register the vector data using the points in common. Generally the vector files are produced by coordinates in plane, which makes transformation easier and diminishes deformations, demanding a smaller number of control points. It may also be necessary to convert several formats of digital vector files to a single format, so that they may be loaded in to the system.

One of the solutions may be the use of the tool *Universal Translator*, in the program *Mapinfo*. This utensil is found in the *Tools* menu and it is capable of converting almost any main vector file format (*.dgn, *.dwg, *.dxf, *.tab, *.mif, *.shp, E00). The command window is divided in two parts – in the first, you specify the type of source file and the location, while in the second you chose the format type and the save location for the converted file. However, the applicative is not free.

Another possible solution is the GPSTrackMaker program. In its free version it is not possible to perform conversions, but it is in its professional version, with low cost, available at www.geostudio.com.br. We accessed this function in File>Convert files: in the first part we chose the source file and in the second, the save format.

We still found the solution based in free programs. It is worth emphasising that, for the LINUX platform there are several solutions that may be utilized for all file formats, like GRASS. For WINDOWS we will use two programs. An important conversions is from *.dxf files to *.shp files. For this procedure we indicate DXFToShapefile, that can be obtained at http://www.freedownloadscenter.com/Best/free-dxf-to-shp.html. For that, install the program by executing the file **DXFtoShapefile_Setup.exe,** click **Next** > until finished, the desktop shortcut **AutoCad DXF to Shapefile Converter** will appear**.** When prompted select the dxf file as a source file format, the alphanumeric save file and the shapefile with the same name and feature type, followed by **Convert**.

More powerful, however, less friendly, is **FWTools**, that can be found at http://www.maptools.org. Download the file for Windows 32 bit, double click file **FWTols221.exe** for install, chose **Next** > **Install** and when finished click **Close**. We recommend you create a folder in a root directory (**C**: or **D**:) and copy all files to be converted to it. Execute the **FWTools Shell** shortcut on the desktop. A small window, similar to the old DOS window, will pop up, where we will perform the conversions. In this command screen allocate the directory where the files to be converted are found. This will be done firstly by observing the directory appearing on the screen, in our example below, "**C**:". Then type **CD\BASE** and press **ENTER**. In case the folder was created in "**D**:", type first **D**:, press **ENTER** and type **CD\BASE**, and then **ENTER**. In case the folder was given another name, replace the word BASE for the folder's name.

In order to perform the conversion, the application follows the following structure: ogr2ogr –f ["type of file to be created"] [name of file to be created] [name of file to be converted]. Example in Figure 2 – converting the file ruas.dgn in to ruas.shp:

ogr2ogr –f "ESRI Shapefile" ruas.shp ruas.dgn, then press ENTER

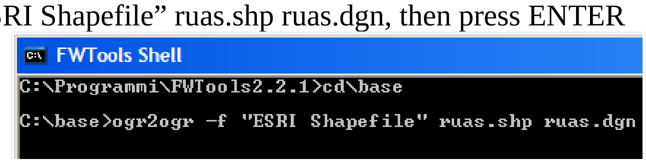

Figure 2. Command line Fwtools.

Remember that the field for the type of file must be written within quote marks, such as "ESRI Shape" or "MapInfo File" and so on. In order to see the main types, type ogr2ogr and press ENTER. At http://www.gdal.org/ogr/ogr_formats.html we may find the complete list of files that can be read and created. We stress the more important, such as: Google Earth .kml = "KML", Arcview .shp = "ESRI Shapefile", MapInfo .mif .tab = "MapInfo File", MicroStation .dgn = "DGN".

c. Table treatment

The main formats for tables are: dbf (as described earlier), xls, mdb, csv and txt (or ASCII), and each of them presents specific characteristics. The dbf files (dbase format) are one of the earlier database formats, however, they are very restrictive regarding the use of accents, symbols, size of field and data type. The xls files (Excel) are read directly by few applications, but the program exports to any format. The mdb files (Access) are a database manager and can contain several tables, are recognizable by several programs and may also export to several formats. The csv files (Comma Separated Value) are a type of file that has been gaining popularity because of being accepted by almost all software and have become a modern format of interchange between lighter databases, since its structure is simple and accepts special characters and symbols. The txt files (text) are also very well accepted, however, when imported, also present problems with characters and symbols.

Even if your system utilises the most modern tools, we still suggest you use the more restricted types of formatting (no special characters or symbols), because you never know to where the data may be transported. We must now begin streamlining the table. The larger restrictions are in the creation of fields. Do not use very long names, spaces, accents, special characters or even symbols.

As for the register, the restrictions are smaller. We may use spaces, dashes, accents, but we should avoid using special characters and symbols. Similarly, regarding the registers, it is of great importance to pay attention to the type of data to be inserted in the table (text, number, etc.). If a numeric field is to be utilized to produces quantitative thematic maps, these must necessarily be numbers. When we use Access, these parameters are defined before inserting the data. However, in Excel, this procedure is automatic and then we may have numeric fields characterized as a text, which may make the use of the data unviable, as when we export them in other formats, they still carry the error.

An important facility of the *Sistemas Gerenciadores de Banco de Dados* (SGBDs) – Database Managing Systems – is the establishing of relationships between table entries, which means saving the data of certain objects in more then one table, facilitating the creation of new tables or complementing between them. In order to be able to use this facility, it is essential that we have common fields between the tables, meaning that this field must be repeated in as many tables as necessary. This is generally accomplished by the method in which the objects are coded or in the use of a common attribute, as shown bellow in Figure 3.

Tabela Bairros

ID	NOME	AREA
B001	Santo Antônio	12,34
B002	Taguatinga	11,67
B003	Lapa	9,60

Tabela Setores Censitários

COD_IBGE	COD_BAIRRO	POPULACAO
314514580	B001	1234
3134095418	B002	234
3134282367	B002	1890

Tabela PSF

UNIDADE	COD_BAIRRO	NUM_AGENTES
PSF011	B001	4
PSF012	B003	5
PSF013	B003	7

Figure 3. Creation of common identification fields.

In this way, for example, we can create a new inquiry for how many health agents we have for the population of a certain neighbourhood, even if this information is in different tables. The inter-operational compatibility between several types of tables and more common programs is simple and may be done with several file formats. From the many applications available in the market, dbf format files are a strongly recommended, even though there are newer formats.The importation of files also presents no great complications. You only need to observe the type of data separator. In the case of txt files, the separator can be the tabulation or the space. On the other hand, on csv files, the separator is the ";",. The dbf files are directly recognized by the program.

d. Table of points

Not uncommonly, we may receive a list of points to be visualized on maps, or we may observe a relation of coordinates in a book or some type of print and have to indicate them on maps.

The vector representation of the point is the most elementary, by it the point may be represented by a single pair of coordinates. In general, it is sufficient to organize them in a table, so that the applications can generate their vector representation. The rules for the field titles are the same as quoted in the section about tables, so to are the observations made regarding file type (extension). However, some directions for the filling of the registers must be given, as seen in Figure 4: minutes

- The coordinates and the abscises must be in separate fields;
- In case of using the "." text formats, the decimals must be separated, and not the ",";
- If using degree coordinates (lat/long), use the maximum of decimals possible. Never use symbols like º, or ". They must be in decimal degrees. Use the sign "+" for North and East and the sign "-" for South and West. As an example, the pair 45º 30'W and 19º 42'S should be expressed as -45.50000 and -19.70000. The transformation can be reached as stated: $DecimalDegrees = Degrees + (Minutes \div 60) + (Seconds \div 3600)$;
- All other attributes must be in separate fields.

COD	NOME	E	N	H
P001	PATO	547234,234	7823567,986	800,000
P002	COBRA	548471,579	7824805,331	837,345
P003	JACARÉ	549708,924	7826042,676	874,690
P004	GALO	550946,269	7827280,021	912,035
P005	MACACO	552183,614	7828517,366	949,380

COD	NOME	LONG	LAT	H
P001	PATO	-42,23456	-18,23567	800,000
P002	COBRA	-42,57956	-18,58067	837,345
P003	JACARÉ	-42,92456	-18,92567	874,690
P004	GALO	-43,26956	-19,27067	912,035
P005	MACACO	-43,61456	-19,61567	949,380

Figure 4. Table of points

3.3. Data acquisition

In cases where data is missing, we must seek other sources or gather the information from the internet or directly in the field with the use of GPS, topography and other similar means. If the data is gathered in the field, with the help of a GPS receiver, we must understand how these may be presented. The equipments called navigation (with no differential correction) produces two types of primitive graphics, the point and the line, since the so called differentiated equipment (with differential correction) produce a list of points in specific software for each brand.

We recommend a study of how the features to be identified in the field will be represented on the SIG beforehand. For example, a building may be represented as a simple point or a polygon indicating the occupied region. Even if we can name the objects recorded, it is important to make complementary notations of important characteristics, thereby building a lasting record of the works undertaken. Since an occurrence may be represented by one or more line or points, it is easy to identify and represent it.

The data from the differential equipment is uploaded in to a program supplied by their makers and result in a list of points that must be worked on adequately. The information from the navigational equipment may be uploaded and treated at the GPS TrackMaker application that presents free and professional versions. The professional version presents the advantage exporting data directly in to SIGs formats. The free version does not presents compatibility with other formats, however, we may resolve this obstacle by exporting the files in the kml extension from *Google Earth* and utilize the FWTools, as already presented.

We recommend using the track memory of the GPS to acquire features such as road axis, school bus roots, street cleaning roots, litter collecting roots, roads and others. This function may be recognized in most devices with the name *Tracks* or *Tracklogs*. They may be collected in a continuous manner and reviewed with GPSTrackMaker. The main alteration would be in the creation of the crossings that may be easily edited with the application. Its important to observe the intersection's distribution, especially in the crossing proximity.

Select all tracks by opening a large window in a way to highlight them by dragging the mouse, then, in the menu **Tools** > **Tracks** > chose **Fragment tracks**. Observe, especially in the crossings, that all roads are now connected. In some cases, these procedures must not be used, like in instances where features with differences in level cross – such as roads, viaducts and trenches – since, in reality, they do not connect. For dot features, like bus stops, residences, schools and health centres, the points are collected normally, we recommend that all crossings be identified with points and that the attributes be noted.

Besides the direct collecting with GPS, the internet has been an important aid in the diffusion of information from cartographic databases and alphanumeric data. However, we must give special attention to the information's quality and their characteristics, in order not to use it inappropriately.
Geominas, an important project developed in the state of Minas Gerais, left a series of digital databases available online as a legacy and as a constant aid, at

http://www.geominas.mg.gov.br/. We recommend the *shapefile* format. The IGAM (Minas Gerais Institute for Water Management) also makes available some databases related to water management data and may be accessed at http://www.igam.mg.gov.br/, follow the links *Geoprocessamento* > Downloads. At a national level, ANA (National Water Agency), at the address http://hidroweb.ana.gov.br/, following the links *Mapas* > *Baixar Bacias*, makes available information on themes such as topography, hydrography, road grids and hydrological basins. No doubt, the main provider of cartographic and alphanumerics data in the country is the *Instituto Brasileiro de Geografia e Estatística* (IBGE) – The Brazilian Institute of Geography and Statistics. During the last two years, the institution has been making available vector data, matrixes and the alphanumeric data from constant gatherings in several areas of activities: ftp://geoftp.ibge.gov.br/mapas/.

Conclusions

We have demonstrated, through the present text, that using only free software and data available on the internet, it is possible to initiate the process of structuring Systems of Geographic Information for municipal managing. The project is in its third edition, and we have already trained 85 municipalities, what means almost 170 people, and until the beginning of 2010 they will be 120 municipalities and more than 250 technicians trained. Besides the classroom courses, the website received a great amount of visitors, searching for the courses handouts, papers and technical support to begin a GIS project applied to urban planning in Brazil.

Special thanks

We thank the Ministry of Cities, which promoted the *Proext-Mec-Cidades* program and supported us in the activities of disseminating geoprocessing in our state (MG-Brasil).

Bibliographic references

GPSTrack Maker. Available in <http://www.geostudio.com.br>. Accessed 10-08-2008.

DXFToShapefile. Available in <http://www.freedownloadscenter.com/Best/free-dxf-to-shp.html>. Access in 10th of August 2008.

FWTools. Available in **<**http://www.maptools.org>. Access in 15th of August 2008.

Geominas. Available in http://www.geominas.mg.gov.br/. Access in 15th of Aug. 2008.

IGAM. Available in <http://www.igam.mg.gov.br/,>. Access in 30th of August 2008.

ANA. Available in <http://hidroweb.ana.gov.br/,>. Access in 30th of August 2008.

IBGE. Available in <ftp://geoftp.ibge.gov.br/mapas/>. Access in 15th of August 2008.

MOURA, Ana Clara M., BORGES, Karla A. V. Avaliação do potencial para implantação do geoprocessamento para a uma prefeitura municipal. Natal-RN, XIV Simpósio Brasileiro de Sensoriamento Remoto, 25 a 30 de abril, 2009.

MOURA, Ana Clara M. Geoprocessamento na gestão e planejamento urbano. Belo Horizonte, Ed. da Autora, 2003. 272 p.

MOURA, Ana Clara M. et all. Geoprocessamento no apoio a políticas do Programa Vila Viva em Belo Horizonte-MG: intervenções em assentamentos urbanos precários. Rio de Janeiro, XXIII Congresso Brasileiro de Cartografia, 21 a 24 de outubro de 2007. p.1544-1553.