A discussion on different techniques for GIS data collecting, precision, accuracy and quality of database

Paulo Cesar Lima Segantine
Engineering School of São Carlos – EESC/USP
Department of Transportation

Rui António Rodrigues Ramos
University of Minho, Engineering School
Department of Civil Engineering

Abstract

Nowadays, is a common sense the importance of geotechnologies in urban planning, transportation engineering and other different areas of knowledge. The Geographic Information System (GIS) is one of the geotechnologies that has been used increasingly. The collection of information and creation of a database are the most expensive, complex and important task in a GIS project. The collection of information results from the direct and indirect measurement of the real world. The reason for creating databases is to register and the maintain the different sources of collecting information. This paper has the objective to present the different techniques for data collection as input in GIS, as well as a brief discussion on the cost associated with the collection of data. Furthermore, comments on precision, accuracy and the quality of database are given.

1. Introduction

We cannot deny the importance of the use of computers and of "computerized information" on the contemporary society. It would be very difficult to quantify man's daily tasks that are in some form related to computational procedures. It is easily agreed that our daily tasks are more and more dependent of the use of computers. As an example, how many times do we get nervous when our e-mail does not work or when the metro delays a few minutes. Computers in recent years are no longer just an important work tool. Now more than ever they are becoming a new way of thinking and working, including in this definition the act of researching, of educating, of communicating etc.

Considering the scientific principle as a basis, it can be affirmed that all types of phenomena are characterized by the spatial location and can be analyzed through processing technologies and teledetection. In other words, can be analyzed through the “geoinformation sciences”. Here, new expressions like Geoprocessing, Geotecnologies, Geomática and Geographic Information System (SIG) are created. These applications will be more and more incorporated into our society’s activities and certainly, the professionals that invest in this knowledge will have great professional advantages compared with those that do not have qualification for their use. It is considered that the geotecnologies will grow in parallel with the increase of capacity and the reduction of the costs of hardware, operating systems and communication.

In view of the increasing importance of the geoinformation for a city, the study and definition of the technique for better data collection, storage and the constant update of the data, becomes necessary, in the aid of the urban planning and environmental protection. The disordered growth of the cities, due to the rural exodus phenomenon that take the people to search for job opportunities and better quality of life, has taken the municipal administrators, and even at a district and national level, to find alternative solutions for the use of geotecnologies in order to assist the decision making process, rapidly, safely and economically. The tools offered by geotecnologies have guaranteed solutions that have been assisting the needs imposed by the population.

Among the geotecnologies, SIG has been a powerful tool and in the last decades, it has become one of the technological fields with the largest growth in developed countries. The use of these systems increases day by day in the support of decision making in several areas, such as environmental, social, economic and politic. This technology offers an excellent tool for the analysis and aid in the administration process, evaluation, monitoring, zoning, planning, territory management, etc.

Currently, almost all GIS are developed aimed at management and making decisions, usually referred to as Support Decision Systems (Longley and Batty, 1996; Birkin et al., 1996; Densham, 1991; Fedra and Reitsma, 1990; Guariso and Werthner, 1989). Therefore, GIS environment dominates the powerful tools for the implementation of space analyses, concerning the continuous evaluation of the territory implemented by the developed model. On the other hand, GIS can also easily allow the preparation of the representative information for a decision-making process, as for example, a slop map, or an inter-visibility map, created from a digital elevation model. Today's GIS allows analysis for all kind of spatial operations. Regarding the combination of data sets from sources at different scales. Not only is it important that the result of such analysis be communicated through well-designed maps, but also that the user be able to judge the quality of the result as well. The question is how to combine different data sources in the best form. This involves a redesign of the measurement strategy, which may affect the data acquisition strategy. Geoinformation needs to be accompanied by information concerning its quality. Without quality, the geoinformation cannot support a reliable decision processes. To improve the accuracy of the geoinformation should be extracted, all errors that eventually occurred during collecting data. This task can be modeled and strategies can be used to allow the corrections. The quality of geoinformation is not only affected by the precision of the equipment, but also the methodologies adopted for data collecting and processing.

2. Geoinformation

Throughout the history, the man always worried about the processing of geographical information for his positioning. The need for a coordinate system that best assists the specificities of his use was noticed along the years. The characterization of an entity with its attributes demanded a complementation for its perfect definition both in space and in time. In older civilizations, the obtaining and generation of geographical information was very limited, probably due to the difficulties related to the collection of data, the graphical representation and the transfer of the information.

The several decisions that involve such concepts as the distance, direction, adjacency, relative location and many other even more complicated spatial concepts are taken regularly in an intuitive form. In this case, man saw himself being forced along the centuries to develop an efficient way of storing information and its complex spatial relations.

When one intends to work with the geoinformation, it is necessary before hand, to consider the use of computers as instruments to represent referenced spatial data, in the attempt to translate the real world into a computational environment. This task has been a lot facilitated by the evolution of the quality of the softwares and hardwares as well as their constants prices reductions, thus becoming of easy acquisition on the part of the different types of users.

The geoinformation is a reality. Its use is growing every year in the most varied branches of sciences, such as social, engineering, mathematical, biological, economics or politics. It is difficult in our days to allow administration, either public or private, to make efficient decision without being aided by the computerized geoinformation.

3. Data collecting

The data acquisition refers to the process of obtaining data in the form that can be inferred with a spatial analysis software of entities and attributes, as for instance a GIS. For a simple level of implementation, this could consist of the ease with which the interpretation of the format of digital dataset supplied by external sources. The effectiveness of the interpretation of this group depends on the ease with which it recognizes the variety of data formats, such as DLG, DXF, SHAPE and NTF, etc., and the "export of formats" of a great variety of programs such as ARC/INFO, ARC/VIEW, IDRISI, MAPINFO, MGE, SPRING etc.

The survey techniques can be used in the primary data acquisition, which is inherent to the areas like Topography, Geology, Aerial or Terrestrial Photogrammetry and Social-Economic Studies. This one, involve interviews and the transcription of documents. The use of GIS facilitates the integration of data collected from different sources, in a way that appears transparent to the final user. The creation of the databases for GIS consists of the cost analysis of the following operations:

- Control of the field survey (Classical Topography or use of Global Positioning System-GPS);
- Aerial Photogrammetry;
- Digital mapping;
- Conversion of existing maps to digital files;
- Selection of data and attributes;
- Digital Ortophotos;
- Remote Sensing;
- Verification and correction of existing data;;
- Verification and correction of data obtained by others sources;

The databases that *feed* a SIG are fundamental for the success of any project. The process of creation of these databases, originating from of any data source, can be divided in two stages. The first one, denominated data input, is understood as the data migration from different sources to digital format; the second, is the conversion of data regarding change in the form of digital representation, because almost always the data should be submitted in some form of conversion to be integrated into the systems.

It is important to emphasize that no GIS can accomplish any useful task until the basic group of space and alphanumeric features has been gathered and acquired. Being thus, a GIS hardly can be become conceived, planned or implanted without previous knowledge of the information that will have to be treated, the present form and the intended use.

The elaboration of database is the most complex and more onerous stage in the implantation of a GIS project. There is no consensus in the bibliography as to what percent of the total cost is spent in this stage. Estimates depend on the objectives of the project, the region and country where its implantation occurs, and mainly of the techniques used for obtaining the database for the project. After consulting some researchers and GIS` agencies, a conclusion that the stage of database acquisition is responsible for the percentage that it varies between 50-80% of the costs and efforts for a GIS implantation.. It is the first stage in the life cycle of a project, and is essential for the success of the whole operation.

What facilitated the incorporation of the GIS technology in our days was the fact that these systems allow the data integration collected from different epochs and scales, also using several different data collection methods.

According to its origin, data can come either from primary or secondary sources. The primary sources are those that directly involve the surveying carried out directly in the field (topographical and geodesic surveying), photogrammetric restitution (terrestrial and aerial) and products of Remote Sensing, etc. The secondary sources are those that refer to the data collected and derived from documents and statistics dataset derived from primary sources, such as graphs, senses, maps, charts, etc.

The task to collect data can be indeed being accomplished through a great variety of techniques and equipments according to the desired application. After the collection of the data, these can be subject to some type of treatment before its effective use. The appropriate use of the data can demand the development of convenient structures of, for instance, insertion, elimination, modification and recovery of registrations. This way, the treatment of data is constituted by a great variety of techniques in permanent evolution.

According to previous affirmation, the collect of spatial data can be done from topographical, geodesic, photogrammetric surveys, by Remote Sensing or still through the transformation of already existing data. The Figure 1 presents some forms of data collection from primary and secondary sources.

The acquisition of the data through those techniques can be done when, for instance, the data is not available or when the existent data are old or not reliable. The transformation of existent data is used when these are reliable, precise, update and appropriate for GIS application.

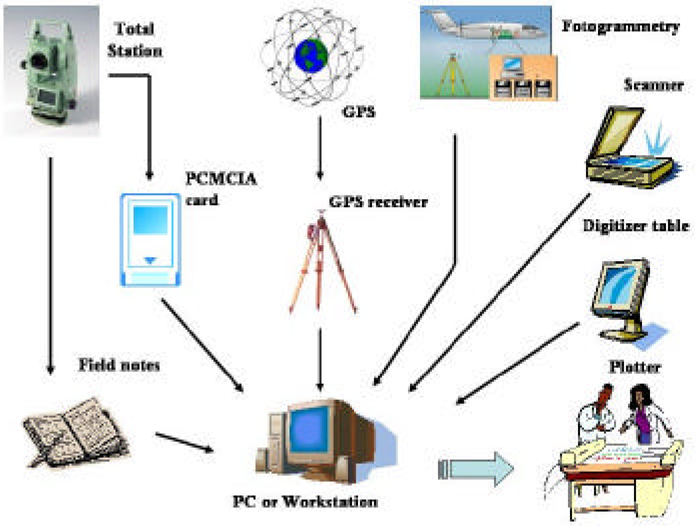

Figure 1: Collect of spatial data.

The source used for the data collecting will depend on the level of desired detailing, expected precision, available resources and the spatial areas. This means that a very detailed planning is necessary in order to obtain what is desired in the implantation of the system. For instance, if the project is in the environmental area, where the precision of the limits (borders) is more flexible, faster and less precise techniques for the collection of data tasks can be used. However, when the project involves a valued urban area, then the determination of limits is essential to get to a good result, depending on more sophisticated techniques for data collection.

The Table 1 highlights the precision of the procedure for data collection, as well as the advantages and disadvantages of its use, and the diverse techniques of collection of data.

Table 1: Techniques to collect data for GIS applications.

Technique	Precision	Advantages	Disadvantages
Manual input data	Not useful	Low cost	Time-consuming (high cost of specialized technicians) no precision on the definition of position.
Total station	< 0,01m	High precision	High cost: need a specialized staff (2 or more technician) and demand visibility between points.
GPS	0,001m – 15m	High precision; fast; the attributes are saved in digital format.	The precision of coordinates can be degraded under top of trees, buildings and reflective surfaces in urban areas.
Photogrammetry	< 0,01m – 1m	Large areas; potential high precision.	High cost by point; limited for attributes.
Satellite image	1m – 1km	Global covering.	Limited for attributes.
Digital maps	0,1m – 100m	Consistent with existent maps.	Need existent maps; the precision is limited by existent products.

From Table 1, it can be affirmed that all the techniques of data collecting are perfectly accessible for GIS applications. What will decide which technique should be used depends on the availability of resources for the capture of the data and the precision intended to give to the project. The duality resources and precision is always presents in the measurements tasks with geoinformation. Financial resources propitiate the acquisition of precise equipments, contract good technicians and the quality of the good project.

The subject of the precision of the landmark limits in urban areas is usually ignored and can cause problems as much as for residents as for utilities. The Figure 2 illustrates two demarcations for the *"freguesias" (lower administrative boundaries)* of Braga Council.

These landmarks were accomplished by different government agencies, with different criteria. One of these demarcation was accomplished by Statistics National Institute (INE) and another used by the Council of the Braga City, based on the 1/25000 scale map.

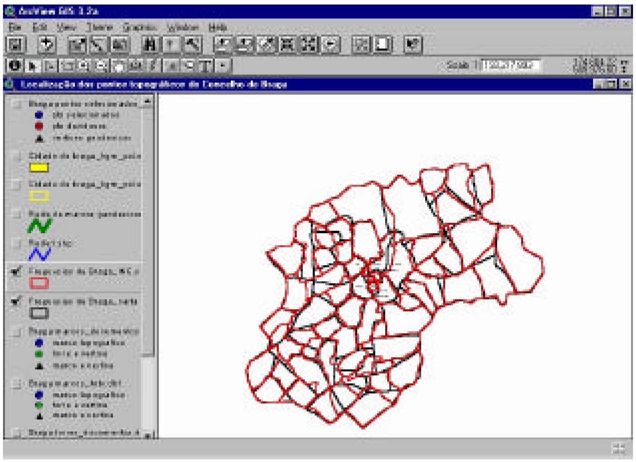

Figure 2: Differences between landmarks of *freguesias*.

This fact could come to bring such problems as a residence that could be in a certain *freguesia* that would not be the correct one. In case where it had differentiation of territorial taxes among *freguesias*, this certainly would become a problem to be solved, because certainly the landowner of this residence would claim from the council his rights. The Figure 3 illustrates an enlargement of the central area of Braga city where the different criteria for *freguesias* delimitation of Braga-São Vicente, Braga-Sé, Braga-São João, Braga-Cividade, Maximinos, São Victor and Real can be observed.

It is important to point out that currently these criteria of *freguesias* delimitation are not applied for definition of property tax collection on the properties. One gives credit that from the moment where there are clear criteria for delimitation of the *freguesias*, certainly, this will be a new form to create contributory differentiation between properties.

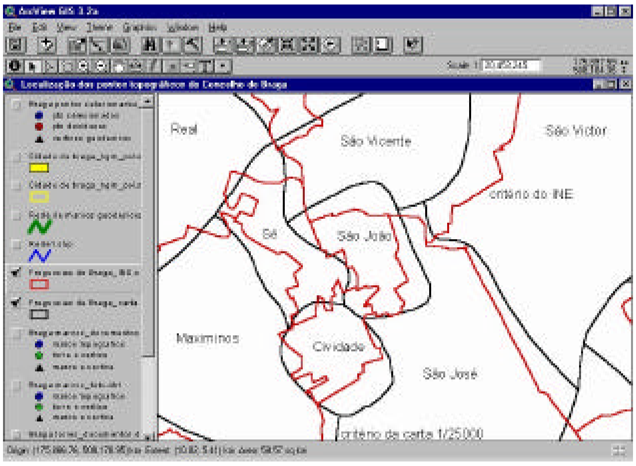

Figure 3: Distinct criteria for *freguesias delimitation.*

4. Precision and accuracy in data quality

With the generalization of the use of electronic measurement equipments, it is common to find different terms that describe the quality of the measures. Amongst these there are two frequently used, but that have their concepts often confused. These terms are *precision* and *accuracy*. The first one, precision, is defined as the refinement degree with which a greatness is measured. In other words, it is how close a series of measurements can be to one another. If a greatness is measured several times, and if the result of these operations is close to one another, then in these cases we can say the precision is high. Usually, the precision is express in standard deviation or the variance of the measurements. Secondly, accuracy refers to the degree of perfection obtained in a measurement, or either, it denotes how close a true value was actually obtained. Just because an operation is of great precision need not necessarily, be an accurate one. The unfamiliarity or the negligence of the conceptual difference between these terms has been causing serious problems in the definition of the data quality, in the location of *freguesias*, as a result in the final cartographic product etc. It is perfectly possible to obtain measurements both precise and accurate, as long as the operator performs the measurements correctly. It is necessary to take precautions while measuring, as for instance, to use checked equipments and to perform correctly all the procedures indicated by the specifics technical norms. It must be in mind that all measurements activities should have the goal to be precise and accurate

5. Quality of geoinformation

When we refer to the concept of geographical information quality is necessary to know that this subject is usually covered by the *quality elements*. These elements are described by the normative documents of the Technical Commission 211 of International Standards Organization (ISO) through ISO/TC 211, 19113 and 19114 and for the Portuguese version of the European Commission of Normalization (CEN) ENV12656. According to MATOS (2001), the relative aspects of the quality of the specified products are not approached in these norms and that should be evaluated in function of the foreseen use and the production cost.

Without fear of committing mistakes, we can affirm that every day exist less economic resources and this fact forces the administrators of public resources to find technical solutions that guarantee the optimization of each cent. The direct consequence of this action is the demand of precise and accurate projects. Happily, the manufacturing companies of mensuration equipments accompanied the technological evolution of recent years and have launched new instruments every year with ever more affordable costs. Three decades ago, the acquisition of good angular measurement instruments incorporating an electronic distance measurement, cost around three to four tens of thousands of dollars. Nowadays it is possible to acquire more compact and better quality instruments for one third of the original value. This fact has contributed greatly to the small survey companies and even city halls of small cities, as they can be equipped with instruments that, since they are used in a convenient way, guarantee quality of the collected information. Another interesting example worth mentioning is the great reduction of the acquisition costs of Global Positioning System (GPS) receivers. Less than a decade ago, these instruments were acquired for amounts greater than hundreds of thousands of dollars and nowadays, after the presence of these instruments in the market, their prices have great amplitude of values. The user can be given the luxury of choosing the color, model, manufacturer etc.

The whole report of the previous paragraph is pure and simply to call the readers attention to always search fast, precise and safe solutions for his projects. Nowadays there are so many norms of procedures that indicate what precision that the project should have or to offer and we have mensuration instruments in the most different levels of precision. It is up to the user to establish the quality that he wants for the project, and from then on take all the necessary precautions given in norms for data collection.

6. Conclusion

This paper did not have the goal to present something extraordinarily new that would to cause an impact in the scientific and academic community. Simply, the objective was to recall the importance of the knowledge of the real notion of the terms precision and accuracy as well as the cares needed in it use. The different techniques that are used in the acquisition of the geoinformation in precise and safe form in order to generate reliable database to be applied in GIS technology is presented. We cannot loose the opportunity to guarantee that without use of computers, the geotecnologies would not have such vast applications and they would not be responsible for so many activities in our actual society. This beginning century indicates that the geotecnologies will be applied more and more in the most different areas of the knowledge due to their potential, and will demand qualified technicians with as the highest possible professional formation to guarantee precise, accurate and safe results.

7. References

Birkin, M.; Graham, C.; Martin, C.; Wilson, A. (1996), *Intelligent GIS - Location decision and strategic planning*. Cambridge: *GeoInformation* International.

Densham, P. J. (1991), Spatial Decision Support Systems. In Maguire, D.J., Goodchild, M.F.; Rhind, D.W. (Eds), *Geographical Information Systems: Principles and Applications*. Harlow: Longman, pp.403-412.

Fedra, K.; Reitsma, R.F. (1990), Decision support and geographical information systems. In Scholten, H.J.; Stillwell, J.C.H. (Eds), *Geographical Information Systems for Urban and Regional Planning*. Dordrecht: Kluwer Academic Plublishers, pp.177-188.

Guariso, G.; Werthner, H. (1989), *Environmental Decision Support Systems*, Chichester: Ellis Harwood.

Longley, Paul; Batty, Michael (1996), *Spatial Analysis: Modelling in a GIS Environment*. Cambridge: GeoInformation International.

Matos, João Luís de (2001). Fundamentos de informação geográfica. *LIDEL – Edições Técnicas, Ltda.* Lisboa, Portugal.

Segantine, Paulo César Lima (2001). Estudo do sinergismo entre os sistemas de informação geográfica e o de posicionamento global. *Tese de Livre-docência. Escola de Engenharia de São Carlos da Universidade de São Paulo*. 230 págs.