QUALITY CONTROL IN
LARGE SPATIAL DATABASES MAINTAINANCE

N. P. Hegde [a,] G.L.Hegde [b] *

[a] Associate Professor, CSE Dept., Vasavi college of Engineering,, Hyderabad, AP, India, nagaratnaph@gmail.com
[b] Project Manager ,Navionics Technologies, Hyderabad, AP,India - gajuhegde@gmail.com

Commission II WG II/7

KEY WORDS: Quality, Spatial Data, Data acceptance Criteria, Automated Quality Control

ABSTRACT:

The heart of any spatial systems is its strong reliable database. The spatial database is a complex in nature which requires special technical skills to maintain it. The discrepancies in database can be minimized by adopting best quality practices for maintaining the database. The systematic approach for updating the large spatial database has been described here. Quality procedure to be followed get the best reliable result has been discussed. Many preventive mechanisms to be adopted to achieve the desired quality has been briefed. Various quality aspects of spatial database such as consistency, completeness, integration etc has been discussed and their mutual relationship has been analysed. The focus is on to achieve the customer oriented quality in spatial database.

1. INTRODUCTION

Data management is a very important and serious business. Data integrity and security are vital because building and maintaining spatial databases are time consuming and expensive and can be central to the core mission of some organizations (e.g., land records administration). Data management often constitutes a large portion of the GIS activities of enterprise GIS organizations.

The database is heart of any spatial system. The functional reliability of the spatial system depends completely on its database. What ever may be the tools used for the other functional aspects of spatial systems, the most important one is the database component.

Geographic databases are becoming a popular subject for research projects. It is acknowledged that database requirements for such applications as Geographic Information Systems (GIS), computer cartography, remote-sensing image databases, and emergency routing/dispatching are distinct from those for traditional database applications; however, the specific database needs due to the properties of geographic data are frequently overlooked. Through the combination of real-time access and geographic data, we identify particularly challenging database requirements of this application with respect to data models, query languages, and query processing

2. QUALITY CONTROL

2.1 Quality in Spatial Data

GIS databases evolve constantly. From paper maps through the digital conversion process to data maintained in a database, GIS data are being constantly transformed. Maintaining the integrity and accuracy of these data through a well-designed quality control (QC) plan that integrates the data conversion and maintenance phases is key in implementing a successful GIS project. Poor data negate the usefulness of the technology. Sophisticated software and advanced hardware cannot accomplish anything without specific, reliable, accurate geographic data. GIS technology requires clean data. Looking at this angle the tools used for the updating and maintaining the database must be latest and most efficient. The process used for maintenance must be simple and well defined and must be understood by all those who involved in the database maintenance . All the guidelines must be clearly given in the process manual. The duties responsibilities of the different functional operators must be clearly written in the process manual. Different quality objectives must be clearly stated in the process manual.

The data maintenance stage of the project life cycle begins once the database has been accepted. GIS data maintenance involves additions, deletions, and updates to the database. These changes must be done in a tightly controlled environment in order to retain the database's integrity. Data are exported out of permanent storage, copied in local storage for update, and then posted back to the permanent storage to complete the update.

2.2 QC Plan review

The fundamentals of quality Control never change. Completeness, validity, logical consistency, physical consistency, referential integrity, and positional accuracy are the cornerstones of the QC plan. All well-designed QC strategies must coexist within the processes that create and maintain the data and must incorporate key elements from the classic QA categories. If QA is not integrated within the GIS project, QA procedures can themselves become sources of error. All the discrepancies noted by the end user of the database must be tracked properly. All these reported discrepancies must be investigated thoroughly if they are found to be valid , appropriate actions must be taken to prevent such errors in future

2.2.1 Visual QC: Creating digital data from paper map sources does not make the data more accurate. Actually this process introduces more or different types of errors into the data. The goal of high-quality data conversion is to limit the amount of random and systematic error introduced into the database.

Random error will always be a part of any data, regardless of form. Random error can be reduced by tight controls and automated procedures for data entry.

Systematic error, on the other hand, must be removed from the data conversion process. A systematic error usually stems from a procedural problem that, once corrected, usually clears up the systematic error problem.

Random and systematic error can be corrected by checking data automatically and visually at various stages in the conversion cycle. A short feedback loop between the quality assurance and conversion teams speeds the correction of these problems.

At various stages in the data conversion process visual QC must be performed. Visually inspecting data can detect systematic errors such as an overall shift in the data caused by an unusually high RMS value as well as random errors such as misspelled text. Visual inspections can detect the presence of an erroneous date, the absence of data, or positional accuracy of data. Visual QC can be performed using hard-copy plots or on-screen views. The hard-copy plotting of data is the best method for checking for missing features, misplaced features, and registration errors. On-screen views are an excellent way to verify that edits to the database were made correctly, but are not a substitute for inspecting plots.

Each error type needs to be evaluated along with the process that created the data in order to determine the appropriate root cause and solution.

Early identification of important and potentially costly errors represents real savings. It is always cheaper to fix a bad process than to correct hundreds or even thousands of errors that may have been introduced into a database. These errors may point to a lack of control during the database update process, may be errors caused by last-minute changes in business rules, could be bugs in the maintenance application, or may be caused by inconsistent editing methods. All these errors can be detected during scheduled validation.

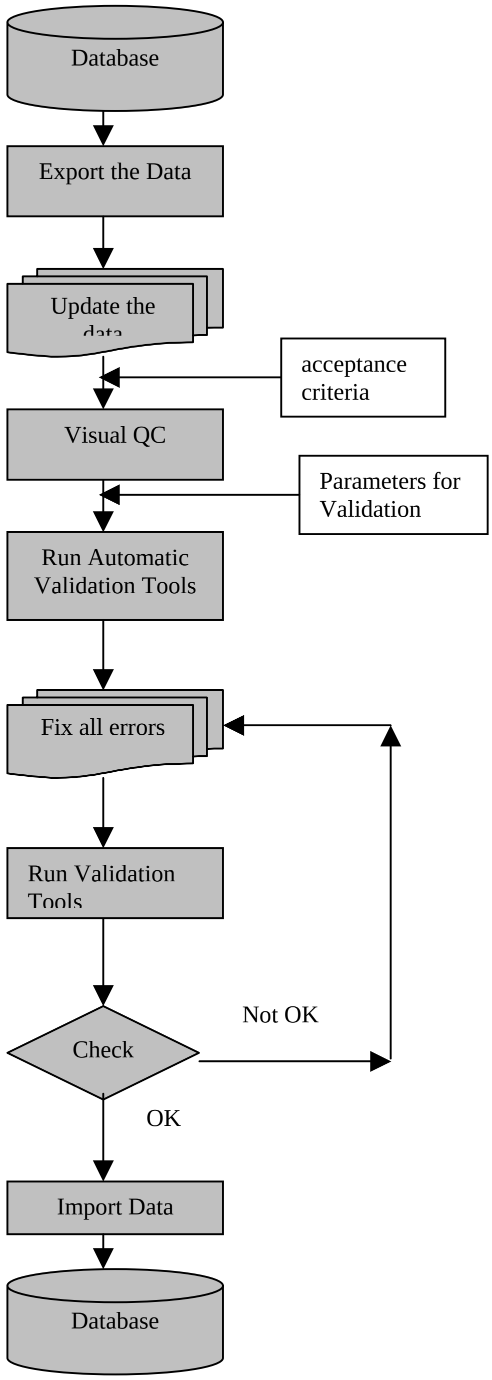

Figure 1.1 Data flow in GIS database update cycle

2.2.2 Edge Matching check : Edge matching, another critical component of the map preparation process, requires that all features that cross or are near the map edge be reviewed with respect to logical and physical consistency requirements as well as noted for positional accuracy and duplication. The temporal factor must be considered. If adjacent maps differ greatly in age there are bound to be edge matching errors between these maps. Cultural features are especially prone to this problem

2.2.3 Conflict Resolution : Conflicts resulting when the same data coming from two or more sources differ must be worked out. Map series must be reviewed together to identify duplicated features and resolve conflicting positional locations and conflicting feature attributes.

2.2.4 Automated QC and Validation : Visual inspection of GIS data is reinforced by automated QA methods. GIS databases can be automatically checked for adherence to database design, attribute accuracy, logical consistency, and referential integrity. Automated QC must occur in conjunction with visual inspection. Automated quality Check allows quick inspection of large amounts of data. It will report inconsistencies in the database that may not appear during the visual inspection process. Both random and systematic errors are detected using automated QC procedures.

2.3 Data acceptance Criteria

All data acceptance criteria must be clearly started in the Process Manual. Which errors are acceptable? Are certain errors weighted differently than others? What percentage of error constitutes a rejection of data? The answers to these questions are not always obvious and require knowledge of the data model and database design as well as the user needs and application requirements. Accepting data can be confusing without strict acceptance rules. Each attribute should be reviewed to determine if it is a critical attribute and then weighted accordingly. Additionally, the cartographic aspect of data acceptance should be considered. A feature's position, rotation, and scaling must also be taken into account when calculating the percentage of error, not just its existence or absence.

2.4 Categories of quality aspects in GIS database

Completeness means the data adhere to the database design. All data must conform to a known standard for topology, table structure, precision, projection, and other data model specific requirements.

Validity measures the attribute accuracy of the database. Each attribute must have a defined domain and range. The domain is the set of all legal values for the attribute. The range is the set of values within which the data must fall.

Logical consistency measures the interaction between the values of two or more functionally related attributes. If the value of one attribute changes, the values of functionally related attributes must also change. For example, in a database in which the attribute SLOPE and the attribute LANDUSE are related, if LANDUSE value is "water," then SLOPE must be 0, as any other value for SLOPE would be illogical.

Physical consistency measures the topological correctness and geographic extent of the database. For example, the requirement that all electrical transformers in an electrical distribution database's GIS have annotation denoting phasing placed within 15 feet of the transformer object is one that describes a physically consistent spatial requirement.

Positional accuracy measures how well each spatial object's position in the database matches reality. Positional error can be introduced in many ways. Incorrect cartographic interpretation, through insufficient densification of vertices in line segments, or digital storage precision inadequacies are just a couple sources of positional inaccuracies. These errors can be random, systematic, and/or cumulative in nature. Positional accuracy must always be qualified because the map is a representation of reality.

3. CONCLUSION

All measurable parameters of the quality must be measured frequently to know the improvements. The measurement data are analysed initially, additional measurements and analyses are conducted as necessary, results are reviewed with relevant users, and necessary revisions for future analyses are noted

Based on data the root causes of the non-conformity, need for corrective action is evaluated to ensure that non-conformities do not recur.
The corrective actions are determined and implemented consisting of

- A definition of the causes of non-conformities.
- Elimination of causes of non-conformities.
- Appropriate actions to avoid recurrence of non-conformities for Corrective action
- Appropriate actions to avoid occurrence of non-conformities for Preventive action

4. REFERENCES

References from Journals:

Gary J Hunter ,"Issues and prospects for Spatial data quality" spatial information technology ICORG vol -1
40-45,BSP publications

Hunter G J, Godchild M F 1995a "a methodology of reporting uncertainty in database products" Journal of urban and regional information systems association vol 7 no 2 11 - 21

Matthew McCain,William C. Masters ,Integrating Quality Assurance Into GIS Project Life Cycle, ESRI ArcNews

Montgomery, Glenn E. and Harold C. Schuch, GIS Data Conversion Handbook, GIS World, Inc. and UGC Consulting, Fort Collins, 1993

References from Other Literature:

Quality management systems – requirements
INTERNATIONAL STANDARD ,ISO 9001: 2000

4.1 Acknowledgements

Authors acknowledge support of Dr. I.V. Murali Krishna, and Dr.KV Chalapati Rao for this article. Authors also thank Cmde. Sudheer Parakala (Chief General Manager Navinics) for his support.