International Archive of Applied Sciences and Technology
Int. Arch. App. Sci. Technol; Vol 10 [1] March 2019 : 148-153
© 2019 Society of Education, India
[ISO9001: 2008 Certified Organization]
www.soeagra.com/iaast.html

IAAST
ONLINE ISSN 2277- 1565
PRINT ISSN 0976 - 4828

CODEN: IAASCA RESEARCH ARTICLE
DOI: .10.15515/iaast.0976-4828.10.1.148153

Geospatial Data Validation Procedure and Techniques

Balakrishnan.M.
GIS and Remote Sensing Researcher, Coimbatore Tamil Nadu.

ABSTRACT
In the past thirty years the requirement to associate geographical information to quality specifications has become significantly evident. Digital geospatial data of various origins and quality are typically integrated in geospatial environments, by determinant associate indefinite level of position accuracy in such systems. Moreover, the introduction of geospatial within the mapping method has created a very new kind of user completely different from the traditional map user. The geospatial data are associate degree ever-evolving entity from their humble early stages as paper maps, through the digital translation method, to the data maintenance stage. The geospatial technology shall comprise of geographic data that's specific and reliable which represents as closely as doable the three-dimensional world, we tend to board and neglecting that, the quality of the technology is temporary. To maximise the standard of geospatial databases there ought to exist a well-designed Quality Assurance set up that's strategically integrated through the whole life cycle of the geospatial development. The quite recently, very little attention was paid to the issues caused by error, inaccuracy and imprecision in the geospatial data sets. This case has modified well in recent years. it's currently usually recognised that error, inaccuracy and imprecision will "make or break" many varieties of geospatial plan. The important purpose is that while error will disrupt geospatial analyses, there are ways that to stay error to a minimum through careful coming up with and ways for estimating its effects on geospatial solutions. Awareness of the matter of error has conjointly had the helpful advantage of creating geospatial expert's additional sensitive to potential limitations of geospatial to achieve impossibly correct and precise solutions. The main purpose of this paper is to alert GIS specialists and its potential users to some strategies that are particularly suited to assessing the standard of geo spatial database and digital maps/ coverage. Tries also are created to gift a group of instructions that are supposed to determine the minimum acceptable level of data quality that must to be followed to by all the users throughout the life-cycle of a Geospatial data.

Keywords: *Geographic Information System, Remote sensing (Spatial Data Quality, Data Completeness, Consistency, Resolution, Thematic Accuracy, Temporal accuracy, Topological inconsistency, Position Accuracy, attribute validation, Strategy of quality assurance level)*

Received 23.12.2018 Revised 11.01.2019 Accepted 14.02.2019

CITATION OF THIS ARTICLE
Balakrishnan.M.Geospatial Data Validation Procedure and Techniques.Int. Arch. App. Sci. Technol; Vol 10 [1] March 2019 : 148-153

INTRODUCTION

Over the last three decades, the utilization of geospatial data and analytics has received substantial attention in business, retail, banking, government and education and other organizations. In additional companies and researchers acknowledge the importance of geospatial information and start to utilize geospatial data to create higher business results. The geospatial data are associate ever-evolving entity from their humble early stages as paper maps, through the digital translation technique to the data maintenance stage. It is now generally recognised that errors, inaccuracies, and inexactitude left unrestrained will build the results of a geospatial analysis practically valueless. Inappropriately, anytime a brand-new data set is loaded, the geospatial additionally inherits its errors. These may be combined and mix with the errors which already in the database in unpredictable ways. The key purpose is that while error will disrupt the geospatial analyses, there are ways that to stay error to a minimum through alert coming up with and ways for estimating its effects on geospatial solutions. Awareness of the problematic of the error has additionally had the helpful good thing about creating geospatial practitioner's additional sensitive to potential limitations of GIS to succeed in impossibly correct and precise solutions.

The important to developing and implementing a flourishing geospatial plan may be a well-designed data validation technic arrange that's integrated with each the information conversion and maintenance phases of the geospatial development. The basics of data validation technics ne'er change for data completeness, topological consistency, physical consistency, referential integrity constraint, and spatial accuracy are the cornerstones of the validation strategy. To maximize the standard of geospatial databases there must be a well-designed data validation arrange that's strategically combined with all aspects of the geospatial plan.

In this paper an effort has been made to make a scientific study of the several quality parameters of a geospatial and their measurements in world atmosphere. This paper additionally presents a collection of strategies that are proposed to determine the minimum acceptable level of accuracy assessment that ought to be followed to by all ofthe developments and users. The most purpose of this paper is to present-day a summary of some strategies that are particularly appropriate to assessing the standard of geospatial information base and digital maps/ coverage. The problems concerned within the development and implementation of an integrated geospatial data validation assurance arrangements are stated.

METHODS AND MATERIALS

The geospatial data validation is commonly used to indicate the superiority of a manufactured good or to indicate a high degree of craftsmanship or artistry. The data quality is a desirable goal achieved through management and control of the production process (statistical quality control). Many of the same issues apply to the quality of GIS databases, since a database is the result of a production process, and the reliability of the process imparts value and utility to the database. For the main purpose of this white paper accuracy of geospatial data is discussed in terms of various methods as Spatial Data Quality, Data Completeness, Consistency, Resolution, Thematic Accuracy, Temporal accuracy, Topological inconsistency, PositionAccuracy, attributevalidation, Strategy of quality assurance level, idea of data maintenance and plan of data acceptance criteria level of the features mapping. The current study will be made for creating or following existing standards and validation procedures will be referenced.

RESULT AND DISCUSSION

Strategy of quality assurance level:

QA is the process of beginning a set of rules or best practices to confirm the production of quality of geospatial data. While performing the quality testing has several benefits like including less data recapture, because quality requirements have formerly been identified and are being measured and monitored continually. QA is usually thought of as acting tests on information to make sure its accuracy. Assuredly, this is QC whereas QC may be a key a part of QA, quality assurance takes a step back and appears at the structure and workflows to implement measures that avoid the introduction of errors. Quality Assurance plans can approximately be classified into two categories. One is Visual QA and another one is Automated QA.

Visual QA:

The Visual inspections will find the existence of error data, missing data or spatial accuracy of information. Visual checks are often achieved by using hard-copy plots or on-screen views. The hard-copy plotting of information is that the best technique for checking for absent feature, incorrect feature, and geo-referencing errors. On-screen views are an exceptional way to verify that edits to the data were created properly however don't seem to be a substitute for inspecting plots. All error type needs be evaluated together with the method that created the information to see the suitable root cause and solution. The Primary identification of significant and probably costly errors represents real savings. It is always cheaper to fix a bad method than to correct thousands of errors that will are introduced into a data.

 ©2019 Society of Education, India

Automated QA:
The Visual validation of geospatial data is reinforced by machine-controlled QA strategies. The geospatial data are often automatically checked for adherence to database design, attribute accuracy, logical consistency, and referential integrity. The machine-controlled QC should occur in combination with visual examination. The goal of the automated quality check permits fast validation of huge amounts of data. it'll report inconsistencies within the data which will not seem throughout the visual check-up method. The both random and systematic errors are detected by Automated QA procedures.

Data acceptance criteria level:
The data acceptance criteria should be clearly started within the Manualmethod. The defining acceptance criteria is probably one of the most troubling segments of the Geospatial project, due to non-availability of Standards for acceptable errors and/or any rejection criteria. Which errors are acceptable? Are certain errors weighted differently than others? What percentage of error constitutes a rejection of data? The answers to those queries aren't continually obvious and need data of the data model and database design moreover because the user needs and application requirements. The acceptive data may be confusing while not strict acceptance rules. A geospatial data set may have 'm' features of 'n' attributes each. Any one feature having a single incorrect attribute, may lead to error-count conditions, such as:

- 1 error, if it does not affect other (n -1) features in any way.
- m errors, if it effects all other features.

Every attribute must to be reviewed to see if it's an essential attribute and so weighted consequently. Also, Additionally, the cartographic aspect of data acceptance should be considered.A feature's position, rotation, and scaling should evenbe taken into considerationonce scheming the share of error, not simply its existence or absence. Once the suitable ratio of error and the coefficient theme are chosen, strategies of error detectionshould to be established. The strategies of error detection for data acceptance are the identical as those employed during the data conversion stage. To check the plots should be compared to the original sources and automated database checking tools should be applied to the delivered data. The very large databases may require random sampling for data acceptance.

Position Accuracy:
The positional accuracy is that the measurable value that embodies the positional change between 2 geospatial layers or between a geospatial layer and reality. Associate in nurture sample of this can be the comparison of the place of roads in a very feature category versus their location in an image/real world. Also, Spatial accuracy is defined as the positional accuracy of features in the geospatial data as measured horizontally and vertically. Position accuracy says that when we see a feature on the map, there's noted probable location at intervals bound limits. For example, if stated accuracy for geospatial (2D) data is +/- 1 Foot, the true location of that feature on the ground will be within 1 foot of the location given on the map. The same applies to vertical accuracy. If elevation data (3D) were stated to be +/- 5 feet accuracy, then the elevation of a feature on the map will be within 5 feet of the true vertical position on the ground. National standards require stated accuracy to be within a 95% confidence level, meaning that of the points tested, 95% fell within the stated accuracy.

 ©2019 Society of Education, India

Figure:1 Positional accuracy as that feature misplacement with real world (OSM Map)

Completeness:
The completeness of data validation refers the presence or absence of features, their attributes, and relationships in a data model. So,the completeness is a validation of the extent and range of a dataset reference to completeness of coverage, classification and verification. The completeness of coverage refers to the quantity of the dataset out there in its entire (including its positional and attribute components). For example, is the spatial data coverage complete for the entire dataset? If not, what quantity of spatial information are incomplete? Attribute data are available for the whole dataset? If not, what quantity of attribute data are incomplete. Consider a dataset of buildings in Bangalore.

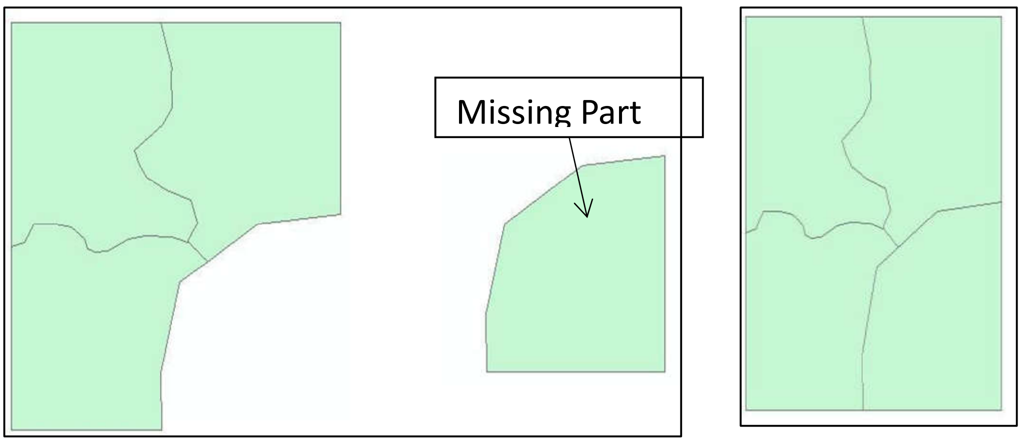

Figure:2 data completeness as that incomplete data as well missing part and complete data.

If it contains solely buildings in Banashankari (one suburban area in Bangalore, instead of all suburbs of Bangalore), it's thought about spatially incomplete. If the dataset contains solely residential buildings, with no industrial, institutional or other private and public buildings, it's thematically incomplete.

Inconsistency:
A spatial data set is alleged to be logically consistent once it complies with the structural characteristics of the particular data model and when it is well-matched with the attribute constraints defined for the set. There are many levels of logical consistency ranging from simple point range or value rules, basic geometric and topological constraints to specific consistency rules for spatial relationships and certain applications.

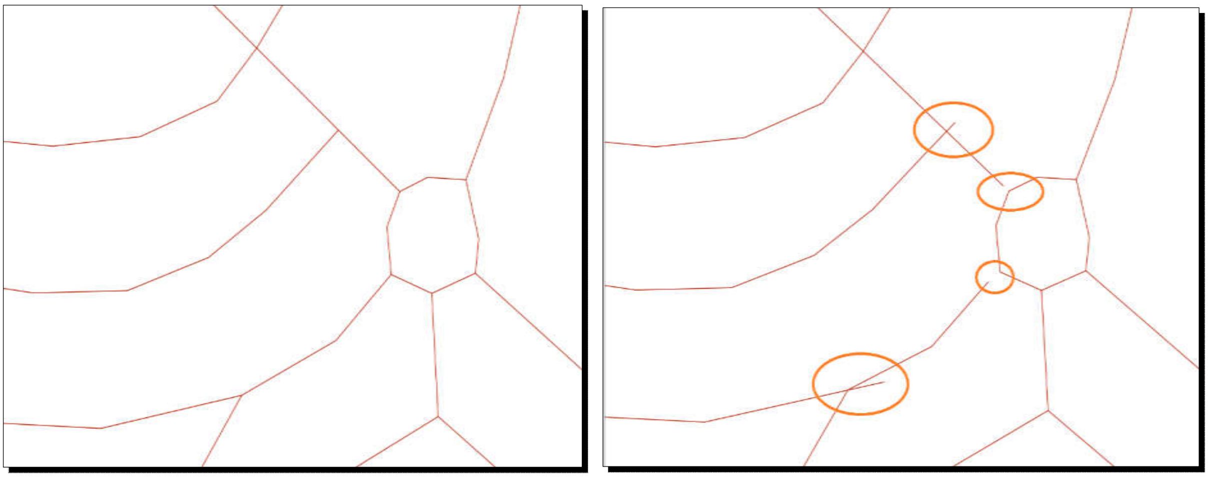

Figure:3 Data consistency as that correct data and error data.

The inconsistencies occur through violating rules and constraints. In some cases, they 'll simply be detected, in others it needs an in-depth analysis and sometimes visual validation to identify rule and constraint violations.

Thematic accuracy:

The thematic accuracy refers to the accuracy of a mapped land use /landcover class at a specific time compared to what was really on the ground at that time. Obviously, to perform a meaningful assessment of accuracy, land use /landcover classifications should be assessed overriding data that are supposed to be accurate. Thus, it's very important to possess a minimum of some knowledge to the accuracy of the referenced data before taking it for comparison against the original maps as remotely sensed data.The below data misclassified the waterbody as highly intensive industry area.

Classified layer *Real world*

Figure:4 Thematic accuracy as that LULC classified layer (error) and real-world data.

Attribute checks:

The attribute checks analyse the attribute values of features and tables. Attribute checking will be utilised for evaluating easy field validation tasks just like a geodatabase domain, or more complex attribute dependencies. The most features usually have one attribute that is reliant on another attribute of the same feature. For example, when a section of a railway track is undergoing maintenance work, it cannot be accessed. An attribute check can be organized to monitor the position and accessibility of the track. Also, the attribute data to validate the field are having empty or Null data and check the attribute data for the same field are having same entry/data. Example: State Name/County Name/District Name field and the check trunk fields as due to the field length restriction some of the attribute get trunked. Example: If name of the filed length has 8 digit/character limit and entering name as "Bangalore". So, the attribute will be trunked as "Bangalor*".

Geospatial - Problem and Solution:

The Geospatial is now a standard technology adopted by the all type of business. The geospatial has replaced their old manual management system that had become costly & inefficient to manage. Initially, many companies implemented GIS purely to simplify the task of maintaining up-to-date records, whereas now, information is recognized as the most valuable resource. The process of maintaining, and updating records, has become extremely essential for a company's continued success within the business. Also, the purpose of data validation is to state the criteria for deciding the degree to which each data item has met its quality specifications as completeness, validity, logical consistency, physical consistency, referential integrity and positional accuracy are essential check to confirm to success the industry.

CONCLUSION

The term validation may be a polite manner of claiming, "Prove it." As geospatial data continues to mix into our decision-making method and enhancing our results with a graphic outcome, the burden of proof weighs heavily. The main aim of this paper is to present an overview of some process that is especially suitable for assessing the standard quality of geospatial database and digital maps/ features. There are many different quality parameters of a geospatial and their measurements in the real world are presented and a set of strategies that are to be adhered to by all the projects and users proposed to establish the minimum acceptable level of accuracy are also highlighted. The standards and references highlighted in this paper are offered as a guide to help you be proactive in protecting yourself and the user from future, costly embarrassments and promote more trust in the use of geospatial data.

REFERENCES

1. Accuracy Evaluation of Ikonos Orthorectified and Stereo Imagery Products, National Imagery and Mapping Agency (NIMA), November 2001.
2. Air Installations Compatible Use Zones, Department of Defense (DoD) Instruction 4165.57, November 8, 1977.
3. Beyond Maps, GIS Decision Making in Local Government, John O'Looney, PhD, Environmental Systems Research Institute, Redlands CA, June 2002.
4. Berry B (1964) : Approaches to regional analysis: A synthesis, Annals, Asoo. American Geographics, 54, 2- 11. (1)
5. Blakmore M (1983) : Generalisation and error in spatial data bases. Cartographica 21, 131- 139. (2)
6. Brassel K. et al, Completeness, 81 – 108 (1)
7. CEN 1995 Geographic Information- Data description- Quality (Draft). Brussels CEN Central Secretariat. (1)
8. Elements of spatial data quality (1995), Oxford, Elsevier eds: Guptill SC & Morrison JL. (6) (2)
9. Geospatial Positional Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (FCDC-STD-007.3-1998), National Spatial Data Infrastructure, Federal Geographic Data Committee, Reston, VA, 1998.
10. Goodchild MF (1991) : Key note Address, Proc. Sympo. On spatial Data base, Accuracy, 1- 16. (3)
11. Kainz W, Logical consistency, 109 -137. (b)
12. Langran G (1992), Time in geographic information systems, London: Taylor & Francis. (3)
13. Redman TC (1992), Data Quality, NY, Bantam. (4)
14. Van Genderen JL & Lock BF (1977) : Testing land-use map accuracy, Photogrammetric enng. & RS, 43, 1135-37. (4) Books/ Standards.
15. Vergin H. et al, An evaluation matrix for geographical data quality, 167 – 188. (c)
16. Map Use, Reading, Analysis, Interpretation, Third Edition, Phillip and Juliana Muehrcke, JP Publications, Madison WI, 1992.
17. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2430044/
18. https://books.google.co.in/books?id=EH5LXJEKLFYC&lpg=PA278&ots=gX_3uMz512&dq=gis%20data%20validation%20conclusion&pg=PA278#v=onepage&q=gis%20data%20validation%20conclusion&f=false
19. http://discoverspatial.in/wp-content/uploads/2018/09/Shahab_Fazal_GIS_BasicsBookFi.pdf

 ©2019 Society of Education, India