

QUALITY CONTROL RELEVANCE ON ACQUISITION OF LARGE SCALE GEOSPATIAL DATA TO URBAN TERRITORIAL MANAGEMENT

A. G. G. Filho^{1,3}, P. Borba³, V. H. S. Silva², A. Cerdeira², A. P. D. Poz¹

¹ UNESP, Science and Technology Faculty, Presidente Prudente - SP, Brazil - (guimaraes.filho, aluir.dal-poz)@unesp.br

² TERRACAP, Brasilia Real Estate Company, Brasília - DF, Brazil - vitor.silva@terracap.df.gov.br, analu.cerdeira@gmail.com

³ DSG, Brazilian Army Geographic Service, Brasília - DF, Brazil - (guimaraesfilho.antonio, borba.philipe)@eb.mil.br

KEY WORDS: urban management, quality control, geospatial data, highest resolution, aerial imagery

ABSTRACT: Quality control (QC) of geospatial data is relevant to urban territorial management to ensure accurate data for government to make strategic decisions when planning cities. The acquisition and control of geospatial data in the Brazilian government must follow INDE - National Data Spatial Infrastructure - through the Technical Specifications. The cadastral cartography from urban areas in Brasília was updated and divided into 10 areas. Acquired data includes classes, features, attributes and metadata on 1:1,000 scale. High resolution images and LIDAR data were used to assist the QC process. The first step of the QC was to check positional accuracy. Samples were applied for each class in the mapping block with 4% rate on the feature random selection and all features class had the same level of confidence. Then, three stages were automatically verified: logical consistency, commission and attribute thematic accuracy evaluations. The process also includes the visual interpretation for omission and classification, which involves a certain subjectivity. Everything was executed with QGIS, FME, Erdas Imagine, Postgresql, PostGIS and a plugin specifically developed for that, the DSGTools. The results show that in general, the quantity of errors were low. However, many errors were detected in the elements completeness and thematic accuracy, specially in áreas 1, 2, 3, 6 and 9. In the opposite, the logical consistency and positional accuracy presented the lowest quantity of errors, which does not diminish the relevance of these errors, since it compromises the usability of the data.

1. INTRODUCTION

1.1. Contextualization

Urban data management is a key method of managing urban public politics, covering a multitude of themes including: land cover, location of public utilities, communication links, sanitation, habitation, conservation and protection of the environment, and historical / cultural heritage. The use of geospatial data is an excellent tool for decision-makers in urban planning as it permits analysis of data from different sectors, thus enabling more informed decisions. Today geospatial information is universally used across a multitude of themes. There are applications that calculate simple equations such as the distance between two locations, how much gas is required to be transferred between different places, as well as property ownership (Meijer, 2015). Demand for geospatial information has grown exponentially with urban development (Jakobsson, 2013). Consequently, the prevalence of multiple sources of data has led to problems with quality of data. Ariza (2018) states that it is necessary to implement more efficient accuracy assessment procedures to obtain fast and easy evaluations of the quality of cartographic products. Jakobsson (2007) introduced two approaches to quality assessment. First, is a procedure that allows the product owner to give recognition to its suppliers to deliver to the required quality, the second approach is a third party attestation that demonstrates the confidence of the quality of the product. Both approaches are important when an official database is created or upgraded. And Albrecht (2018) points out the role of standards in support of validation services.

1.2. Data Quality - ISO

The ISO 19157:2013 establishes the principles for describing the quality of geographic data. From that, many countries created specific standards and guidelines to manipulate Spatial

data. Besides, these standards are a part of a spatial infrastructure, e. g. INSPIRE (European Union, 2007) and IDE-CHILE (Gonzalez, Neira and Riquelme, 2014).

The quality of spatial data, or products derived from it, has the following elements or components: Completeness (C), Logical Consistency (LC), Positional Accuracy (PA), Temporal Consistency (TC), Thematic Accuracy (TA) and Usability element. All of these elements are called Internal Quality Elements of a dataset, besides, there is external quality, but won't be discussed in this paper.

Thus, technical guidelines create parameters, values and specifications under the influence of ISO standards, documents, laws and decisions. In the case of Brazil, the National Data Spatial Infrastructure - INDE (Brazil, 2008) was created in 2008 and has many components, among which there are the Technical Specification for Geospatial Quality Data Control - CQDG (DSG, 2015) and for Structuring of Vector Geospatial Data - EDGV (CONCAR, 2008).

1.3. INDE - Brazil

According to Warnest (2005), a Spatial Data Infrastructure has five components: People, Data, Institucional, Technology, Standards and Specifications. And they are strongly related and interact among them.

INDE was conceived to allow data sharing between Brazilian government agencies, brazilian GIS industry and the society. To add to data interchange, a definition of a basic data structure to standardise its exchange was required. For this purpose ET-EDGV (portuguese acronym that stands for Technical Specification for Structuring of Vector Geospatial Data) was created. This specification defines a conceptual data modelling structure for vector data produced by Brazilian government.

1.4. Spatial Data Quality - Brazil Technical Specification

Under the influence of ISO 19157, ISO 2859-1:1999, ISO 2859-2:1985 and others national technical standards, the CQDG (portuguese acronym that stands for Technical Specification for control of quality for Geospatial Data) was created to assign quality to a geospatial production environment to control the quality of geospatial products that are purchased from third parties. This requires the advent of principles, components and concepts of quality, including sampling procedures. These characteristics are key to the quality of Brazilian Government Official Geospatial Products. The aim is to use concepts of ISO in Brazilian geospatial infrastructure, e.g. products, methods, and resources.

2. METHODOLOGY

2.1. Area

The city of Brasilia is a big city, known as Brazil capital. But Brasilia is one part of Federal District State, with more than 2 million people, and the last urban land mapping was at 1997 in 1:2,000 scale. To update Geospatial database of urban area, the government contracted a mapping services of a private company to work on 1,200 km², split into 10 areas (Figure 1). Each area has an urbanization rate, population density and specific features.

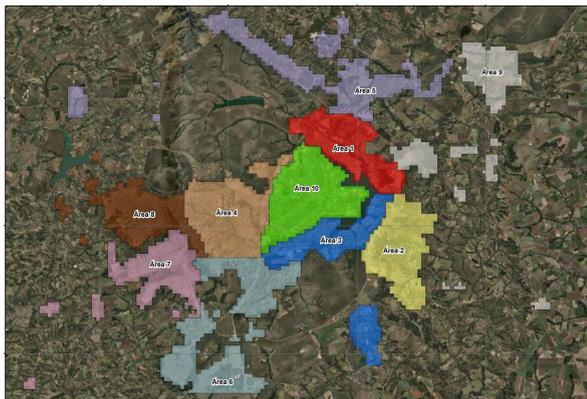


Figure 1. Urban area split into 10 mapping blocks.

2.2. Spatial Data

In this work, there are two types of spatial data. One is the geospatial database which contains classes, features, attributes and metadata (acquired in 1:1,000 scale) that will be analysed and the other would be used in Quality control. For that, a very high resolution images and LIDAR data (density of 4 pts per m²) are the auxiliary data for Quality Control (QC) process.

The reference orthoimages for evaluation have following specifications: Ground Sample Distance (GSD) = 10cm ; RGB bands, 8 bits per band, SIRGAS 2000 geodetic reference system, UTM 23S cartographic projection. These data were obtained by airborne imagery in 2016 and it was a reference to acquisition geospatial database. In addition, other auxiliary data: surveying data, like a road and river names, house and parcel numbers, as well as preservation areas.

2.3. Methods

Before the implementation of workflow, it is necessary to organize and standardize which elements and sub-elements will be checked and the manner (automatic or visual), Table 1 shows general scheme for quality control.

(QE)	Subelement	(QC)	Obs
(C)	Omission	yes	Visual process with orthoimages
	Commission	yes	Automatic Process with minimum size area and width line
(LC)	Conceptual Domain	not(1)	-
	Format	yes	Postgresql PostGIS
	Topological	yes	Automatic Process to verify spatial relationships defined by EDGV.
(PA)	Absolute or external	yes	Manual process. The operator select a sample points of features to check accuracy with orthoimages and surveilling data.
	Relative or internal accuracy	not	-
	Gridded data	not(2)	-
(TA)	Classification correctness	yes	Automatic process to Attributes with default values. Visual process to classification of elements.
	Quantitative attribute	yes	
	Non-quantitative attribute	yes	

Table 1. Elements and Subelements Quality. QE: Quality Element.

(1) the database was created by QGIS plugin DSGTools. It was done to avoid implementation errors, which it would lead to logical consistency errors;

(2) the positional accuracy of orthoimages are compatible with scale 1:1000, spatial resolution between 4 to 21 cm and coordinate error below 28 cm with standard deviation up to 17 cm.

In this work a workflow was created to organize the sequence of process. The main idea (fig. 2) is to concentrate visual analysis in few stages, experts analyse complex tasks (like thematic accuracy), to save information on shared database to all team, as well as to have a review stage. Moreover, another important aspect is the standardization of analysis process and report, avoiding subjectivity in the data evaluation.

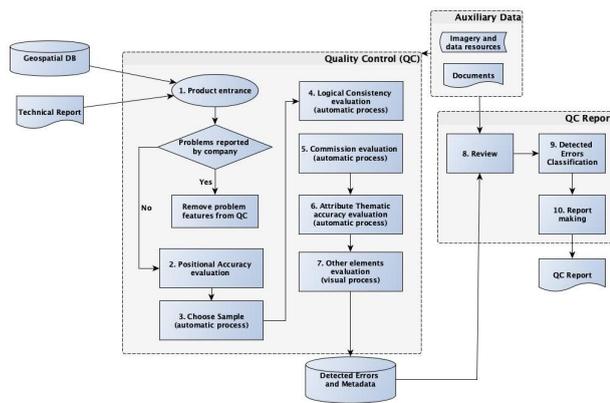


Figure 2. Main workflow to (QC)

According to figure 2 there is a stage called “problem reported by company”. The reason for that is some constraints that happens during acquisition of geodata, for example: violence areas, private areas without authorization and bad weather conditions. All features with these constraints don’t follow main workflow, only logical consistency and positional accuracy were checked on separate process.

The QC has 10 steps split into two phases, geospatial database and technical reports as inputs, QC report as output. Besides there is the auxiliary data (imagery, data resources, general documents, e.g., specifications, laws, etc.) as reference.

First step is checking positional accuracy (PA). Evaluation of altimetry and planimetry is basic for all geodatabase and the CQDG defines horizontal error below 28 cm with standard deviation up to 17 cm, vertical error below 50 cm with standard deviation up to 17 cm, both cases to 90% points at 1:1,000 scale map.

Another important step is choosing the samples that were applied for each class in the mapping block with 4% rate on the feature random selection. All features class had the same level of confidence in the QC. After those, there are three automatic stages: logical consistency, commission and attribute thematic accuracy evaluations. The software used in this step are described in section 2.4. Besides, in case of attribute evaluation, it was filled in the standardized way. In other words, all attributes were filled by codelist and not handwritten in database. This decision make possible the automatization and decreased bundlers.

Visual process is a part that demands more attention, mainly because involves image interpretation. Omission and classification are subelements of quality with a certain subjectivity and some requirements were settled down. For example, normally QC occurs after the entrance of product, however in this project some doubts (e.g. specifications on technical guidance, thematic features classification, etc.) were responded before the data acquisition. As well as during this process, questions and doubts were responded. Consequently, the subjectivity of classification would be reduced. At the end, there are three steps focused at report. The idea is to create an objective report that a third party could to understand as well as to execute the necessary changes.

2.4. Project Software Stack

The software stack used in the scope of the aforementioned project is QGIS 2.18.14 (QGIS Development team, 2016),

FME Desktop (FME, 2018), Erdas Image (Erdas, 2014), Postgresql 9.5 (Postgresql, 1997) and PostGIS 2.2.1 (Refractions Research Inc, 2005). Natively QGIS did not fulfill all the Brazilian Army Geographical Service's needs, so a plugin DSGTools was developed to meet those requirements (Andrade et al., 2016). It has many geoprocessing tools, some of them developed to quality assurance and quality control of geospatial data. Erdas Image was used for stereoscopic viewing of images, PostgreSQL and PostGIS as GIS database storage backend and FME to integrate data, to test or filter data and to report results. DSGTools is a QGIS plugin written in Python 2.7. It is a geospatial suite that enables data production using PostgreSQL. The processes used to assess the quality is similar to the one described by Borba et al. 2018: Remove Empty Geometry, Deaggregate Geometries, Adjust Coordinate Precision, Identify Invalid Geometries, Identify Not Simple Geometries, Identify Vertex Near Edges, Identify Duplicate Geometries, Identify Small Areas and Identify Small Polygons, Identify Dangles, Identity Gaps, Identify Overlaps, Identify Gaps And Overlaps in Earth Coverage Polygons, Spatial Relator.

3. RESULTS AND ANALYSES

In summary, Table 2 shows the quantity of features that were checked (samples) and total features of mapping blocks. These results were compiled from thirty reports made in all QC process.

area	(TF)	(FS)	%(C)	%(LC)	%(PA)	%(TA)
1	166273	4420	58.3	0	0.05	12.71
2	187679	1482	21.52	1.48	1.21	2.70
3	152249	3090	17.80	0.03	0.06	6.38
4	106164	1377	5.23	0	0.58	1.96
5	226984	1580	7.91	0.32	0	1.90
6	401731	5284	6.96	0.02	0.08	9.41
7	308137	3795	7.27	0.03	0.16	4.30
8	412060	4355	3.81	0	0.05	3.38
9	318762	4331	31.45	0.05	0.18	3.86
10	227543	4464	2.37	0.16	0.02	1.50

Table 2. Compiled QC Results. TF: Total Features; FS: Feature Sample.

The results showed in Table 2 were obtained by first version of geospatial database. Only third version of geodatabase was approved with quality requirements. In general, the quantity of errors were low. However, in case of completeness and Thematic accuracy, many errors were detected, mainly in areas 1,2,3,6 and 9. Since Area 1 was the first mapping block delivered, many errors were found and these errors were fixed when other blocks were mapped. However, systematic errors were detected in completeness, mostly omission, and thematic classification. The elements Logical Consistency and Positional accuracy presented the lowest values of errors, nevertheless it is still a big problem. Because futures analysis will be

compromised and the coordinates of features don't have confidence, in other words, the database will be useless to the aim proposed. Some error cases detected by quality control are below (Figure 3). In the quality report is important to explain the shown error, classify and quantify it.

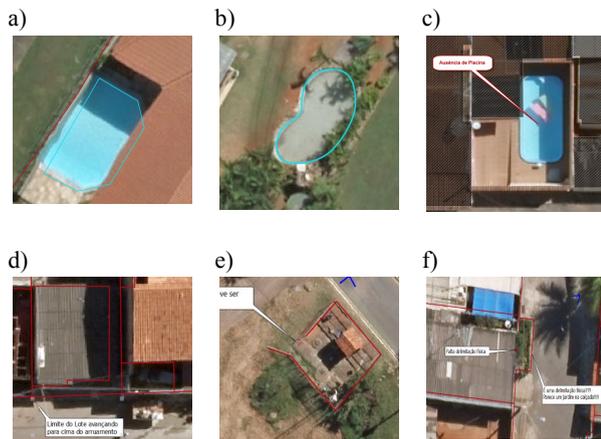


Figure 3. Error examples: a) pool acquisition (positional accuracy); b) pool acquisition (thematic accuracy); c) missing pool (completeness omission); d) wall and light pole (positional and thematic accuracy); e) fence and gate (positional accuracy); f) wall (completeness omission) and garden (thematic accuracy)

There are many reasons for that, for example acquisition teams with different training, misunderstood about some classes (definition), problems with translating features, feature edition with invalid geometry, and others. At the end, all mapping blocks were delivered without logical consistency and positional accuracy problems, with completeness and thematic accuracy up to 1% tolerance error. The final dataset is available on site <<https://www.geoport.al.seduh.df.gov.br/>>.

4. CONCLUSION

This paper showed the relevance of data quality control on geospatial data. Even the best method to data acquisition has errors. The aim of quality control is to analyse inconsistencies regarding aspects of tolerance values and specific rules so that products or data can be useful for the intended purpose. Without QC, none analysis can be done or if were, the results will be wrong and it could impact negatively on decisions. After the acceptance of mapping blocks, the local government ensure that the product or data to new urbanism projects, land regularization, land registration or cadastre and upgrade property taxes. A Geospatial data with quality is essential to planning and management of cities, to make spatial distribution of population and economic activities fairer, avoiding and correcting distortions at urban growing, negative effects on the environment and to promote the economic and social prosperity.

ACKNOWLEDGEMENTS

The authors thank Brasilia Real Estate Company - TERRACAP for sharing reports and results of Quality Control.

REFERENCES

Albrecht, F., Blaschke, T., Lang, S., Abdulmutalib, H. M., Szabó, G., Barsi, Á., ... Huang, G. (2018). Providing Data Quality Information For Remote Sensing Applications. ISPRS - International Archives of the Photogrammetry, Remote Sensing

and Spatial Information Sciences, XLII-3, 15–22. <https://doi.org/10.5194/isprs-archives-xxii-3-15-2018>

Andrade, L. C. O.; Borba, P.; De Paulo, M. C. M. Agregação Hierárquica de Geo-Objetos: Uma Abordagem Baseada em Software Livre. Revista Brasileira de Cartografia (2016), Nº 68/8: 1531-1540.

Ariza-López, F., Ruiz-Lendínez, J., & Ureña-Cámara, M. (2018). Influence of Sample Size on Automatic Positional Accuracy Assessment Methods for Urban Areas. ISPRS International Journal of Geo-Information, 7(6), 200. <https://doi.org/10.3390/ijgi7060200>

Brazil. Decree no 6.666, of November 27th, 2008. Establishes within the Federal Executive, the National Spatial Data Infrastructure – NSDI (Considered interesting to complement). Official Diary of Brazil's Federal Government, Brasília, DF, Nov 28th, 2008. Section 1, p. 57.

Borba, P., Esperidião, J.P.O.S., Filho, A.G.G., Silva, G.H.O. Controle de Qualidade de Redes Unidirecionais Topologicamente Conectadas Utilizando Software Livre. VII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação Recife - PE, 08-09 de Nov de 2018.

CONCAR, National Commission of Cartography. Technical Specification for Structuring Geospatial Vector Data – ET-EDGV. Version 2.1.3. Brasília – DF: MPOG, 2008.

DSG, Brazilian Army Geographic Service. Technical Specification for Geospatial Quality Data Control – ET-CQDG. 1a Edition. Brasília – DF: DSG, 2015.

ERDAS (2014) ERDAS Imagine 2014. Hexagon Geospatial, Peachtree Corners Circle Norcross, US.

European Union, 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Published in the official Journal on the 25th April, 2007.

FME (2018) FME Desktop 2018. Safe Software Inc, 137A St Surrey, BC Canada.

Gonzalez, C., Neira, J. and Riquelme, C. (2014). Basic Territorial Information Group Spatial Data Infrastructure, SDI-CHILE (SNIT). In: FIG Congress 2014. [online] Kuala Lumpur, Malaysia: International Federation of Surveyors (FIG), p.7 pages. Available at:< https://www.fig.net/resources/proceedings/fig_proceedings/fig2014/papers/ts04h/TS04H_gonzalez_neira_et_al_6802.pdf>.

International Organization for Standardization, 1999, ISO 2859-1:1999: Sampling procedures for inspection by attributes — Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection.

International Organization for Standardization, 1985, ISO 2859-2:1985: Sampling procedures for inspection by attributes — Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection

International Organization for Standardization, 2013, ISO 19157: 2013: Geographic Information—Data Quality.

This contribution has been peer-reviewed.

Jakobsson A., Giversen, J., eds., 2007. Guideline for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies. Eurogeographics Expert Group on Quality.

Jakobsson, A., Hopfstock, A., Beare, M., & Patrucco, R. (2013). Quality Management of Reference Geo-Information. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-2/W1, 127–132. <https://doi.org/10.5194/isprsarchives-xl-2-w1-127-2013>

Meijer, M., Vullings, L. A. E., Bulens, J. D., Rip, F. I., Boss, M., Hazeu, G., & Storm, M. (2015). SPATIAL DATA QUALITY AND A WORKFLOW TOOL. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-3/W3, 55–59. <https://doi.org/10.5194/isprsarchives-xl-3-w3-55-2015>

PostgreSQL. Califórnia, Estados Unidos da América, 1997. Disponível em: <<http://www.postgresql.org>>. Acesso: 22 de julho de 2018.

QGIS Development Team, 2016. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>

Refractions Research Inc. PostGIS. Canada, 2005. Disponível em: <<http://postgis.net>>. Acesso: 22 de julho de 2018.

Warnest, M. A collaboration model for national spatial data infrastructure in federated countries. 2005. Dissertation (Ph.D. in Geomatics)–Department of Geomatics. University of Melbourne, Australia,2005.