Assessing the Quality of Cadastral Geodata Administrated by the Municipality of the City of Los Ángeles, Chile

Guido Staub*, Francisco Almendras*

*University of Concepción, Department for Geodetic sciences and Geomatics, J. A. Coloma 0201, 4451032 Los Ángeles, Chile

Abstract. Cadastral information is sparsely available in Chile. Only a few municipalities handle relevant information and in such a way that it really is a benefit for the community in terms of urban planning, sustainable development, land management and so on. The municipality of Los Angeles is among the few municipalities that intend to implement a cadastral information system. Almost up to date aerial photographs and vector data of road networks, city blocks, park areas, etc. are already available. Nevertheless, little or almost nothing is known about the quality of this data, as it has never been checked in an appropriate manner using for example spatial statistics. Neither exists a standardized update procedure at regular basis to assure timeliness of the geodata. Therefore, standardized quality assessment is carried out using statistical analysis as well as visual and in-situ comparison of the available geodata. This study reveals that not only positional accuracy is weak; displacements of up to 3 meters were observed. Also attribute accuracy is not sufficient for the implementation of a cadastral information system, as no standardized catalog for geodata description and storage exists. Temporal accuracy, data completeness and logical consistency are well for recently build city blocks and would require only some minor modifications. Nevertheless, in case of city blocks that were built a couple of years ago, in particular temporal accuracy and data completeness are weak. In conclusion, since all the mentioned quality aspects are of importance for future tasks that permit to enhance quality of life, we propose several measures, like e.g. usage of international standards for geodata description and storage, coordinate transformation. Such measures would allow to improve, enrich and to complement the available data, without the necessity to replace all the available information by geodata collection and surveying.

Keywords: Cadastre, Data Quality, Surveying

1. Introduction

In Chile, cartographic information at national scale is elaborated by the military geographic institute of Chile. By law, the regular cartography is generated at 1:50,000 scale and is used as the basis for small scale topographic maps coverage at scales like 1:250,000 and 1:500.000. Furthermore, a significant number of cities also count with urban digital cartography at larger scales (1:1000 to 1:5000) as Barriga & Advis (2002) stated out.

The cartographic information of the city of Los Ángeles was developed using different techniques and methodologies since it's initial stage in 1995, when airborne photogrammetric observations (1:20000) were used to extract the cadastral core geodata. The original cartographic database was corrected and improved with the comprehensive plan for evacuation and drainage of rainwater in Los Ángeles (1:2500) and the restitution of eleven road axis from the Urban Transport System analysis report (1:1000). In 2003, geometric information on land parcels was obtained from an aerophotogrametric restitution (1:1000). Since then, continuous, but not regular, updates were incorporated based on topographic surveys in areas of new urbanization.

As a result and due to the applied methodology in the past, geodata quality has deteriorated. The scale of cartography for the urban area is 1:1000. When working at a greater graphic scale, for example for analysis purposes, significant geometric errors are perceptible.

The main objective of this study is to determine the quality of the cadastral data administrated by the municipality of the city of Los Ángeles. This is in particular of importance because several tasks, like urban planning, transport, security, water and waste water management and energy supply, directly depend on geodata that comply with geodata quality measures.

To address the quality of geodata, accuracy and precision are commonly used, although they are semantically different and often misused in geospatial literature (van Oort, 2006). A high quality dataset is very often understood as a conjunction of data with high geometric precision, low uncertainty or error. Gervais et al. (2009) declare accuracy as the degree to which map data or a geodatabase matches reference data and/or it's values, while precision refers to the level of measurement and correctness of information in a geodatabase. Arsanjani et al. (2013) discriminate internal and external geodata quality. Internal quality, as defined by e.g. ISO (International Organization for Standardization), ICS (International Classification for Standards) or FGDC (Federal Geographic Data Community), is composed of attribute accuracy, positional accuracy,

temporal accuracy, logical consistency and completeness (Guptill & Morrison, 1995). These data properties refer to data production standards and specifications that are based on existing errors in the data and are saved in metadata files (Devillers et al., 2007). External quality considers suitability of a dataset for a particular purpose and whether it meets or exceeds a users expectation of information (Devillers & Jeansoulin, 2006).

In Campbell & Shin (2011) several measures by which data quality can be described and determined are mentioned:

- Positional Accuracy: typically expressed by the Root Mean Square Error (RMSE) to evaluate the difference between a measured feature and either its true location (absolute positional accuracy) or its location in relation to other mapped features (relative positional accuracy).

- Attribute Accuracy: expresses to which degree values of a recorded feature are correct and not missing within the attribute field.

- Temporal Accuracy: refers to the age or timeliness of a dataset.

- Logical Consistency: allows to make a statement about the topological correctness of geodata.

- Data Completeness: evaluates whether all features to ensure accurate mapping results are present within the GIS database or not.

- Semantic Accuracy: describes whether a feature representation is complete or not. Both object type and feature attributes have to been given.

2. Methodology for Quality Assessment

2.1. General Aspects

Positional accuracy: The study of this quality criterion was carried using the approach described by the Federal Geographic Data Committee (FGDC) (NSDI, 1998). First of all a minimum of 20 check points has to be measured and the difference between the obtained coordinates and the corresponding coordinates from the cartographic information has to be calculated. Afterwards the Root Mean Square Error (RMSE) in X and Y direction has to

be determined. Finally the positional accuracy has to be estimated using a confidence interval of 95%.

Attribute accuracy: This measure has to be evaluated for each object type individually. As there is no attribute catalog established for data storage, It cannot be determined whether a specific field is missing or not. Only completeness can be checked and established fields can be reviewed for missing or incorrect values.

Temporal accuracy: This accuracy criterion has to be evaluated by a review of the time stamps of each contribution made to the geodatabase. As there is no plan for regular updates established, temporal accuracy differs from dataset to dataset and from block to block.

Logical consistency: Although, a quantitative indicator to addresses this measure does not exist, it is possible to carry out a visual evaluation. Polyline data such as roads are analyzed to detect whether or not road junctions fulfill topological consistency, beginning and ending of road segments have to match. Point data as it is used as for tree representation can be checked as well; have to be placed within green corridor or front gardens.

Data Completeness: Another aspect that has to be checked visually as it highly depends on temporal accuracy. It is meant to describe the degree of data absence in a dataset to a reference dataset. For relative data completeness estimation, a comparison between two or more datasets has to be carried out (the one that has the highest temporal accuracy of all the available datasets has to be considered as reference). Furthermore a visual inspection outdoors to estimate absolute data completeness can be realized.

Semantic accuracy: It describes how good geographical entities in terms of their types and their attributes and semantics are represented by the geodata. In our study, we check whether object type and its attributes are given or not. If it was like this, the object has the maximum semantic accuracy.

2.2. Test case – City of Los Ángeles

For positional accuracy estimation, we realized fieldwork and measured checkpoints. Statistical analyses lead to RMSE determination for the area of interest. It is expressed in meters and indicates whether a spatial displacement tendency exists or not.

The results of the other quality measures are expressed verbal. Our description reaches from *Very Low* over *Low*, *Intermediate* and *High* to *Very High*. Very low is assigned in case that only 15% of the studied features comply with the specific measure. Low for 16% to 35%; Intermediate for 36% to 65%; High for 66% to 85%; Very High 86% to 100%. Features that we checked in this study are trees, road networks, lots and structures on these lots.

Figure 1. *City of Los Ángeles and location of the 4 neighborhoods analyzed in this study (marked by red polylines)*

3. Material

The present study was performed in four particular neighborhoods of the city of Los Ángeles as shown in Figure 1, namely Villa Galilea, Villa Parque Sor Vicenta, Villa Jardines de la República and Villa Balmaceda. They have been chosen because there is almost nothing known about the geodata quality and its cartographic accuracy. In the past, the information was

acquired little by little as relevant information became available and constructions were completed.

3.1. Aerial Photographs

During January 2012 an aerial photogrammetric survey was realized. The municipality of Los Ángeles acquired a total of 180 aerial photographs captured with a 50 Mega Pixel Hasselblad H2 camera. The geometric and radiometric quality of the photographs is high due to the characteristics of the camera. So far they have not been processed any further in a photogrammetric manner, rather planimetric corrections were applied based on significant ground control points. Therefore we decided to orthorectify the aerial photographs, to guarantee their geometrical correctness.

3.2. Vector Data

The Municipality of Los Ángeles has a huge data collection of a variety of relevant cadastral information stored as ESRI Shapefiles. These files are basically used for planning and decision support. Although all of them are geo-referenced they lack on a uniform reference system. This mainly relies on the fact that in Chile a legal framework for cadastral information does not exist. Therefore, for example construction companies are not obliged to present development plans in a uniform geodetic reference system. In consequence they opt for local reference systems.

3.3. GNSS Observations

As it is proposed by the FGDC, for positional accuracy analysis, a minimum of 20 check points have to be established and measured. These points have to be distributed regularly, covering the geographic area of interest. The measurement of the checkpoints was carried out using a single frequency GPS receiver in Stop & Go mode. Post processing accuracy can be estimated to 1cm +/- 1ppm.

	Villa Galilea	Villa Parque Sor Vicenta	Villa Jardines de la República	Villa Balmaceda
Number of check points used	32	42	26	27

Table 1. *Overview of number of check points measured and used for analysis*

Check Point selection was basically driven by the possibility to detect them in terrain and in the aerial photographies. Therefore most of them have been placed at fences or perpendicular road intersections. The number of

check points measured in the already mentioned neighborhoods and finally used in this study can be found in Table 1.

4. Results

Our results reveal some important facts about cadastral data quality for the studied city blocks in the neighborhoods of the city of Los Ángeles. The data sets of the four city blocks that were analyzed in this study are almost complete. Positional accuracy tends to be better in north as in east direction. For Villa Balmaceda the measures do not achieve a satisfying result, as all quality measures are worse. This is based principally on two reasons. First, the heterogeneity of Villa Balmaceda, see figure 2. Different types of constructions, inner courtyards and gardens can be identified in the aerial Photographies, which was confirmed outdoors. Several abandoned sites were detected but do not appear with a correct semantic within the geodata. Furthermore, cadastral information about trees is incomplete and logically incorrect. And second, Villa Balmaceda is the oldest neighborhood. Cadastral information source for vector data generation was elaborated once and since then it was never checked or updated. Any modification has never be registerd by the municipality.

Figure 2. *Study area Villa Balmaceda*

In case of Villa Galilea (see figure 3), Villa Parque Sor Vicenta and Villa Jardines de la República the results indicate that cadastral information has high quality in almost all the analyzed aspects. They are more homogeneous as Villa Balmaceda and data collection, elaboration and description were carried out with more diligence.

Figure 3. *Study area Villa Galilea*

Table 2 highlights the results of our quality assessment study, carried out as described in section 2.1, for the four neighborhoods in the city of Los Ángeles.

	Villa Galilea	Villa Parque Sor Vicenta	Villa Jardines de la República	Villa Balmaceda
Positional accuracy	N: 2,067m E: 3,145m	N: 0,775m E: 2,307m	N: 2,248m E: 2,882m	N: 0,430m E: 2,146m
RMSE (Radial)	2,433m	3,655m	2,188m	3,764m
Attribute accuracy	Intermediate	Intermediate	Intermediate	Low
Temporal accuracy	Very High	Very High	Very High	Intermediate
Logical consistency	Very High	Very High	Very High	Low

Data completeness	Very High	High	Very High	Intermediate
Semantic Accuracy	Very High	High	Very High	Intermediate

Table 2. *Result of data quality assessment*

5. Conclusions and Outlook

With our results it is possible to conclude that the cadastral information managed by the municipality of Los Ángeles needs to be restructured and revised in terms of positional accuracy. A directional tendency can be determined, which indicates that a displacement exists between the vector data and the real world. The reasons for this issue can be found in the different data sources, the geodata delivery and elaboration in different geodetic reference system and the missing legal framework for geodata presentation.

Attribute accuracy in general can be determined as intermediate and as it does not reach high nor very high, it is questionable that in its current description is sufficient for cadastral information.

Almost all features are present, although modernizations might result in a lack of information as seen at Villa Balmaceda. Completeness, semantic and temporal accuracy depend on initial data acquisition and on regular updates.

Logical consistency is given in case of Villa Galilea, Villa Parque Sor Vicenta and Villa Jardines de la República. The reason is that they are homogeneous city blocks and until now have not suffered any major modifications. For example, road networks are taken from original plans are not exposed to changes.

All the mentioned results and conclusions have to be considered for future studies, which might have the aim to implement a cadastral information system. As a standardized feature catalog for data description and storage is missing, such a catalog has to be envisaged. It should be possible to make use of IDE (Infraestructura de Datos Geoespaciales) Chile[1], a governmental approach for the implementation of a unified geodata environment. In case of the municipality of Los Ángeles, some important improvement can be achieved applying simple measures like coordinate transformation (6 parameters affine transformation, tested by Almendras (2014)) to enhance

[1] IDE Chile: http://www.ide.cl/

positional accuracy, establish a compulsive update scheme, complete metadata and so on. Nevertheless, at national scale, this is not an issue that can be solved at municipal scale. Definitely, a nationwide legal framework has to be established. Therefore, we recommend to design and implement a centralized data management system that complies with national or international standards to assure data quality.

References

Almendras, F. (2014)Control Posicional de la Cartografía Catastral de la Ciudad de Los Ángeles. Thesis, 91 pages

Arsanjani J, Barron C, Bakillah M, Helbich M (2013) Assessing the Quality of OpenStreetMap Contributors together with their Contributions. Proceeding of AGILE 2013.

Barriga R, Advis M (2002) Experience odd the Military Geographic Institute of Chile in the national infrastructure for ego spatial data. GSDI 6 Budapest Hungary 2002

Campbell J, Shin M (2011) Essentials of Geographic Information Systems. Saylor Foundation

Devillers R, Jeansoulin, R (2006) Fundamentals of Spatial Data Quality, ISTE, London

Devillers R, Bédard Y, Jeansoulin R, Moulin B (2007) Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data, International Journal of Geographical Information Sciences, vol. 21, no. 3, pages 261–282

Gervais M, Bedard Y, Levesque M, Bernier E, Devillers R (2009) Data Quality Issues and Geographic Knowledge Discovery, in Geographic Data Mining and Knowledge Discovery, second edition, Miller & Han (Eds.), pp. 99–116

Guptill S, Morrison J (1995) Elements of Spatial Data Quality, Elsevier Science, New York

NSDI (1998) Geospatial Positioning Accuracy Standards. Part 3: National Standard for Spatial Data Accuracy. U.S. Geological Survey. FGDC-STD-007.3-1998

Van Oort P (2006) Spatial Data Quality: From Description to Application (PhD Thesis), Wageningen University, the Netherlands