http://dx.doi.org/10.15576/GLL/2015.1.69

GLL

Geomatics, Landmanagement and Landscape No. 1 • 2015, 69–75

BDOT500 DATABASE OF PHYSICAL TOPOGRAPHIC OBJECTS – BASIC QUALITATIVE ANALYSIS

Marek Ślusarski

Summary

Official databases, which gather spatial data, should include sets of metadata, which serve to describe the information within. The fundamental element of a metadata set consists in the features describing the quality and relative importance of geospatial data. In the present work, we propose a method for the evaluation of database quality pertaining to databases of topographic objects (BDOT500) based on four criteria: location accuracy, completeness, validity (in the sense of being up to date) and logical coherence. Overall quality evaluation of the BDOT500 database was carried out using the calculation of two estimation factors: average absolute value and the coefficient of average variation. The method described herein is a useful device, which allows for a quick and credible evaluation of the BDOT500 data quality at a basic level. Applying the method does not require involving any additional means, as all the necessary information is already recorded inside the database.

Keywords

topographical object database • metadata • spatial data quality

1. Introduction

Information technology revolution in the second half of the twentieth century has led the world into the era of information, in which the resources of collected data are maintained by computerised information systems. These systems enable simple data sharing, as well as performing complex analyses in order to offer processed data. After a short period of belief that computer databases contain reliable information, research was started upon the quality of IT data, broadly understood.

According to Redman [2001] "data is of high quality when it is suitable for the use in operational, decision making and planning processes". Features of high quality data include: availability, coherence, and usefulness. Data should be equipped with appropriate imprints (data sheets). These imprints should be clear, measurable, easy to obtain, and they should facilitate comparability of results.

Compendium of geospatial data infrastructure, The SDI Cookbook [Nebert 2004], distinguishes discovery metadata, which makes it possible for the user to assess the quality of the data set and determine the data of the set according to his needs. The main elements of the CSDGM standard (US Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata) include (in the order of importance) [Longley et al. 2006]: basic information on the data set, information on the data quality (general assessment of the quality of data in the set), manner of organisation, or ordering of geospatial data in the set, and more.

While describing the quality of geospatial data, different features or properties of data can be used. Data origin, precision of location, precision of attributes, logical coherence, completeness, semantic precision, and temporal quality are the main elements of data quality [Oort 2005, Devillers 2010].

According to Gaździcki [2008], the quality of data is described with the following features: completeness, logical coherence, positioning accuracy, temporal accuracy, thematic accuracy, as well as semantic accuracy and derivation. Completeness is understood as the incidence of all the intended data, without deficiency but also without excess. Logical coherence is the lack of inner contradictions in the data set. Positioning accuracy pertains to geodesic accuracy – the location of objects, expressed with coordinates. Temporal accuracy is linked to the change in data over time, and thematic accuracy, to the correctness of determining the qualitative properties, for instance. Semantic accuracy presented as a data set reproduces the area of discussion (that is to say, the subject matter of the problem). Derivation describes the manner and time of obtaining the data as well as the source materials, methods, and techniques.

European norms of the ISO 19100 series contain a wide range of terms pertaining to geographical information, and they possess a rich conceptual framework. The comprehensive description of the quality of data is included in the ISO Norm 19157 "Geographic information – Data quality" [ISO Norm 2013]. According to the record of the norm, quality equals "overall product characteristics, which depend on its capacity to meet particular and potential requirements". A complete identification of data should include both "non-quantitative" and "quantitative" information on quality. Non-quantitative information includes: purpose, origin, and application. Quantitative information includes, among others: completeness, logical coherence, location accuracy, temporal accuracy and thematic accuracy. Completeness is an element of data quality, which examines the deficiency and/or excess of the data set. Logical coherence is considered on four levels: conceptual, pertaining to the given area (domain), topological, and related to the format. Location accuracy is either absolute or external, or relative or internal. Temporal accuracy is the accuracy of time measurement, temporal coherence and temporal relevance. Thematic accuracy examines the correctness of classification, the precision of non-quantitative attributes, and the precision of quantitative attributes.

2. Subject of the research

Currently, for the area of all Poland, databases are maintained covering sets of geospatial data pertaining to the spatial data infrastructure. The main databases include:

- register of surveying grids / geodesic matrices (PROG, BDSOG),
- land and building register (EGiB),
- geodesic infrastructural network register (GESUT),
- national register of borders (PRG),
- real estate prices and value register (RCiWN),
- register of topographic objects (BDOT10K),
- register of geographic names (RNG),
- register of photographic images (RZF),
- database of topographic objects (BDOT500).

The database of topographic objects (BDOT500) – according to the provisions of the regulation on geodesic database for the infrastructural network register, database of topographic data and the master map [Regulation no. 383/2013] – is a database of topographic objects with the degree of detail which informs the creation of standard cartographic elaborations on the scale of 1 : 500 through to 1 : 5000. The content of the database includes objects of spatial infrastructure, which are not covered by the land and building register (EGiB) or the geodesic infrastructural network register (GESUT). The objects have been grouped in three levels of classification. The first level features categories of object classes, the second, the classes of the objects, and the third, the objects themselves.

Classes of the objects included in the BDOT500 are:

- engineering structures (bridges, overpasses, etc.),
- hydraulic engineering structures (weirs, dams, etc.),
- sports structures (tribunes, running tracks, tennis courts, etc.),
- tall technical structures (cooling towers, chimneys, etc.),
- technical tanks and reservoirs
- road, railroad or waterways reinforcements (groynes, retaining walls, etc.),
- earthworks (slopes, dikes, etc.),
- transport equipment (cranes, lifts, etc.)
- streets, tracks and surface waters,
- natural objects (trees, waterfalls, etc.).

BDOT500 database is established on the basis of the materials from the state surveying and cartographic resources. Basic sources of data include archival technical geodesic reports and assessments (operates), the master maps and other large-scale maps.

Empirical studies and experimental investigations of the BDOT500 data quality were conducted using a test object. The town selected for the study was of average size (town with administrative rights of a district, i.e. poviat) located in the Region of Silesia. It is an industrial town with over 62 thousand inhabitants, covering the area of approximately 64 square kilometres. Official databases of geospatial data are kept for the town area in the digital format. The BDOT500 database was established chiefly on the basis of the master map and geodesic measurement reports (operates). The IT system supporting the BDOT500 database facilitates, among other functions, the updating of the data, cartographic visualisation of the data, and data provision (that is, making it available).

3. Research methodology

An empirical study has been conducted: an analysis of the information contained in the BDOT500 database. Information describing the objects in the following classes has been analysed: engineering structures, hydraulic engineering structures, sports structures, tall technical structures and earthworks. Assessment of the quality of the data included in BDOT500 was carried out based on four properties: location accuracy, completeness, relevance (up-to-date quality) and logical coherence. Assessment results for each particular property have been described as point values in the scale of 1 to 100.

Data quality for the property "location accuracy" was calculated based on data source attributes determined for each particular object in the BDOT500 database. For the studied object, there are three types of data sources. Data source marked with the symbol *O* designates the location of the object, determined on the foundation of geodesic measurements performed based on the geodesic network (Table 1); data source marked with the symbol *F* designates the measurement based on photogrammetrical surveying; while data source marked with the symbol *D* denotes the digitalisation of the master map. Point values for the assessment of location accuracy were calculated through analysing the numbers of objects which posses the attributes of *O*, *F* and *D* relative to the total number of objects.

Data sources were assigned measures of accuracy and weights (of relative importance). Subsequently, weighed values of points were assigned for the feature of "location accuracy" (Table 1).

Table 1. Calculation of point values of the data quality assessment for the property of "location accuracy"

Symbol of data source	Measure of accuracy [m]	Weight	Points	Weighted points
O	0.15	1.00	38.6	38.6
F	0.25	0.36	9.3	3.3
D	0.45	0.11	52.1	5.8
Total			100.0	47.7

Source: author's study

Completeness is the criterion, which determines the ratio of the number of collected data entries to the number of data entries, which should be collected according to the assumed theoretical model [ESDIN 2010]. For the studied databases, completeness of data was calculated based on the analysis of objects of the BDOT500 type, which exist only in the digital master map. These objects have not been registered in the BDOT500 database, mostly due to the lack of complete identification.

Relevance, or the up-to-date quality of data depends on the character of data (its variability) and the assumed method of its updating. Ensuring a high level of relevance requires the application of on-going updates, therefore, entering updated information to the BDOT500 database, each time the new information or information change occurs. Analysis of waiting times (delay in entering the new or changed information into the database of geodesic reports) allowed for determining the value of the "data relevance" criterion.

Logical coherence is defined as the lack of contradictory relations within the data. For BDOT500, what needs to be checked is collinearity and the correctness of line objects segmentation [Bielecka 2010]. For the analysed database, conditions of collinearity and continuity of objects have been checked as well as the correctness of line objects segmentation. Research was conducted using the tools of the QGIS software.

Logical coherence of the database has been graded very highly, due to the conformity with pertinent legal regulations and technical norms when establishing the BDOT500 database. Table 2 presents scores (point values) of data quality assessment for the four characteristics, described using point values in the scale of 1 to 100.

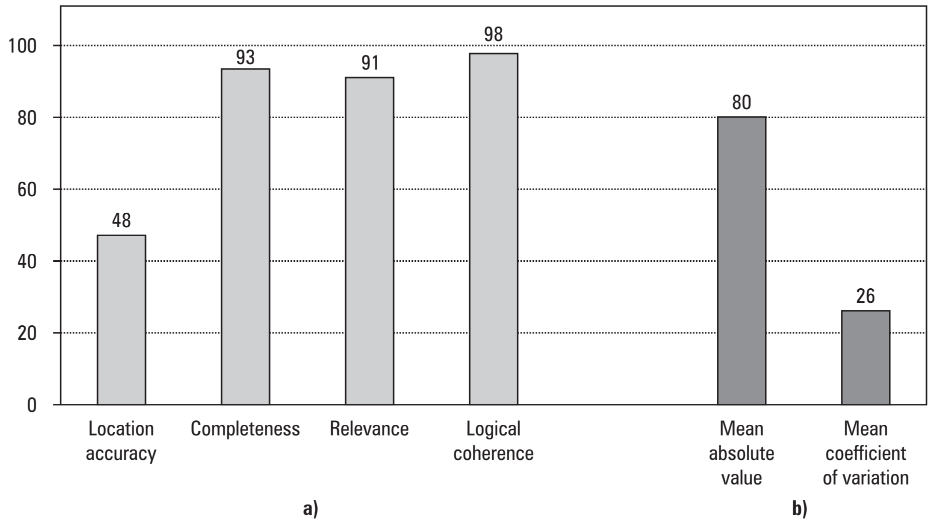
Overall estimation of the BDOT500 database quality was conducted using the calculation of two estimators: mean absolute value (*Sa*) and mean coefficient of variation (Λ). The values of estimators were calculated based on point assessment of the data quality criteria as well as assumed weights (Table 2). Coefficient *Sa* – the mean of all data quality criteria – indicates which portion of the data meets the assumed criterion of 100 points. Parameter Λ facilitates a comparison of variation in different distributions [Czaja 2001] – in this case, the BDOT500 database which is heterogeneous in terms of data quality. In the ideal model, mean coefficient of variation model equals zero.

Table 2. Point values of the data quality assessment and calculated values of estimators

Criteria of data quality				Assessment estimator	
Location accuracy	Completeness	Relevance	Logical coherence	Mean absolute value (*Sa*)	Mean coefficient of variation (Λ)
weight = 1	weight = 1	weight = 1	weight = 0.5		
48	93	91	98	80	26

Source: author's study

Point values of data quality assessment presented in Table 2 and in Figure 1 indicate that the BDOT500 database of the examined object displays a decidedly average value of the location accuracy parameter (48 points). Other quality criteria – completeness, relevance, and logical coherence – display high values (above 91 points). Mean absolute value for the given object scores at 80 points, while the mean coefficient of variation (expressed as percentage) is 26. Values of assessment estimators demonstrate, that the conceptual model of the BDOT500 database has been implemented to a high degree of completeness. However, not all aspects impacting the quality of the collected data have been sufficiently taken into account – which is particularly visible in the case of location accuracy.



Source: author's study

Fig. 1. Point values of the data quality assessment (a), values of estimators (b)

4. Conclusions

Official databases, which gather spatial data, should include sets of metadata, which serve for the description of the information within, in such a way as to facilitate the determination of data usefulness. The fundamental element of a metadata set consists in the features describing the quality and relative importance of geospatial data.

The method for the evaluation of database quality pertaining to databases of topographic objects (BDOT500), proposed in the present work, is based on four criteria: location accuracy, completeness, validity (in the sense of being up to date) and logical coherence. Empirical studies and experimental investigations of the test object have shown that the four parameters allow for a quick and credible evaluation of the BDOT500 data quality at a basic level. Applying the method does not require involving any additional means, as all the necessary information is recorded inside the database.

BDOT500 database of the test object is a set of reliable geospatial data. The data quality criteria – completeness, relevance, and logical coherence – meet the stated criteria in over 90%. Only the criterion of location accuracy is not met in approximately

50% of the analysed cases. The reason lies in the fact that the analogue master map was the main resource for the establishment of the BDOT500 database. Very high overall data quality score of the examined BDOT500 database encourages the use of that set as base material for creating other geospatial databases, for instance the BDOT10k register of topographic objects.

References

Czaja J. 2001. Metody szacowania wartości rynkowej i katastralnej nieruchomości. KOMP-System, Kraków.

Bielecka E. 2010. Zasady oceny danych przestrzennych oraz ich zastosowanie do oceny jakości danych gromadzonych w TBD, Roczn. Geomat., VIII, 4, 40, 53–66.

Devillers R. (ed.) 2010. Fundamentals of spatial data quality. Wiley-ISTE, Hoboken, 178–197.

ESDIN. 2010. D 8.4 ESDIN Quality Final Report – Part A. http://www.esdin.eu/ sites/esdin.eu/files/D8–4_ESDIN_Quality_Final_Report.pdf, 17–29.

Gaździcki J. 2006. Leksykon geomatyczny, http://www.ptip.org.pl (accessed: 10.02.2015).

Longley P. et al. (2006). GIS teoria i praktyka. PWN, Warszawa.

Nebert D. (ed.) 2004. The Spatial Data Infrastructure Cookbook, http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf (accessed: 10.02.2014).

Norma ISO 19157. 2013. Geographic information – Data quality measures. Geneva: International Organization for Standarization.

Oort P. 2005. Spatial data quality: from description to application, Delf: NCG, 13–18.

Redman T. 2001. Data quality. The Field Guide. Digital Press, Boston, 233.

Rozporządzenie Ministra Administracji i Cyfryzacji z 21 marca 2013 r. w sprawie bazy danych geodezyjnej ewidencji sieci uzbrojenia terenu, bazy danych obiektów topograficznych oraz mapy zasadniczej.

Dr inż. Marek Ślusarski
Uniwersytet Rolniczy w Krakowie
Katedra Geodezji
31-198 Kraków, ul. Balicka 253a
e-mail: rmslusar@cyfronet.pl