http://dx.doi.org/10.15576/GLL/2023.4.235

GLL

Geomatics, Landmanagement and Landscape No. 4 • 2023, 235–245

HARMONISATION OF THE EGIB, GESUT AND BDOT500 CONCEPTUAL MODELS WITH THE BDOT10K DATABASE IN TERMS OF DATA QUALITY AND SUPPLY

Marek Ślusarski

Summary

The main source of supply in the process of creating and updating the BDOT10k database is data from the official EGIB, GESUT, and BDOT500 databases. For this reason, the degree of harmonisation of these datasets is crucial. The study of the level of harmonisation of the EGIB, GESUT, and BDOT500 datasets with the BDOT10k database in terms of data quality and supply was performed by analysing the conceptual models of the discussed spatial data registers. The study of the degree of harmonisation of datasets was performed by classifying the analysed object classes into one of three groups, characterising the consistency of the attribute values: full, partial, and none. Analysing the degree of harmonisation of the conceptual models in terms of spatial data quality, it was found that the possibility of feeding quality properties from the EGIB, GESUT, and BDOT500 databases into BDOT10k is limited. BDOT10k has definitely richer characteristics regarding data quality. Analysing the extent of harmonisation of conceptual models in terms of feeding the data of the BDOT10k database with data from EGIB, GESUT, and BDOT500, significant inconsistencies were found. The data supply from these databases is limited especially in terms of descriptive attributes. Acquisition of geometric axes of linear and surface objects does not pose major technical problems. Potential difficulties will arise from, among others, topological and geometrical errors, such as close proximity of vertices or different segmentation. Correcting the conceptual model of the EGiB, GESUT and BDOT500 databases in order to achieve a high degree of harmonisation with the BDOT10k model is feasible. All that is needed are appropriate institutional formal and legal actions.

Keywords

spatial databases • data harmonisation • data integrity

1. Introduction

Access to reliable and up-to-date spatial information and the possibility of using this information in decision-making processes is a key problem in Poland, as well as in other European countries. One of the ways of solving this problem is through measures taken by public institutions and the economic sector, aimed at the exchange, use, access, and exploitation of both spatial data and the services concerning these data. Such activities make it possible to establish national spatial data infrastructures. With such infrastructures, the interoperability of datasets, metadata, and spatial data services can be achieved [Baranowski 2012]. Interoperability is defined as the ability to combine datasets and service interactions without manual intervention, in such a way that the result is consistent and the added value of the datasets and services is enhanced [Directive 2007].

The process adopted to reach the interoperability of the components of the spatial information infrastructure is harmonisation. The Geodetic and Cartographic Law [2023] defines the term harmonisation of datasets as legal, technical, and organisational measures aimed at the mutual consistency of these datasets and their adaptation for joint combination and use. Harmonisation of official spatial datasets improves, among other things, the quality of spatial information, avoids multiplication of the same data by different institutions, and facilitates the feeding and exchange of data between scattered databases.

The land and building register (EGiB), geodetic register of utilities network (GESUT), and database of topographic objects (BDOT500) are official databases collecting data characterised by the highest level of detail. According to the Geodetic and Cartographic Law [2023], the EGiB database is defined as an information system ensuring collection, updating, and making available, uniformly for the entire country, information about land, buildings, and premises, their owners and other entities that own or manage these lands, buildings or premises. The currently functioning EGiB databases are created and updated in accordance with the provisions of the Land and Building Register Ordinance of 2021.

The GESUT database – as stipulated in the Geodetic Law – is defined as an information system that ensures the collection, updating, and making available the information on utilities networks, uniformly for the entire country. The utility networks are all kinds of above-ground, ground, and underground conduits and devices, as well as underground structures. The currently functioning GESUT databases are created and updated in accordance with the provisions of the Regulation on the geodetic register of utilities network of 2021.

The BDOT500 database – as stipulated in the Regulation on the database of topographic objects and the base map of 2021 – is a database of topographic objects, detailed enough to ensure the creation of standard cartographic studies at scales of 1:500–1:5000. The contents of this database are spatial development objects, which are not collected in the EGiB and GESUT databases. Objects in the BDOT500 database include engineering structures (bridges, viaducts), hydrotechnical structures (weirs, dams), streets, tracks, and surface waters.

The relationship between the BDOT10k and BDOT500 conceptual models has previously been assessed by other authors. Lupa et al. [2018], who analysed the feasibility of feeding and updating the BDOT10k with the BDOT500 data, found that there are significant differences due to the purpose and level of conceptual generalisation. The BDOT500 database is less conceptually accurate than the BDOT10k. There are significantly more attribute features in the BDOT10k. Over half of the object classes have various types of inconsistencies. More than 30% of the common classes have no data exchange between registers. Only 3% of the common classes show full compatibility of conceptual models.

With regard to feeding the BDOT10k database with linear and surface objects derived from EGIB, GESUT, and BDOT500, the applied methods of simplifying the geometry of cartographic objects are of particular importance. The development of methods for generalising (simplifying) the geometry of objects in digital cartography mostly involves improving the existing algorithms without validating them in terms of the similarity of the geometry of objects before and after the process. Geometry generalisation also relies on evaluating the results of the algorithms, i.e. the features needed for automatic generalisation. The preparation of fully automatic generalisation of spatial data requires the use of certain standards along with unique and verifiable algorithms for specific feature groups [Kronenfeld et al. 2020].

2. Materials and methods

EGiB, GESUT and BDOT500 are official databases that collect data of the highest level of detail and are therefore used to feed other systems in the process of their creation and updating. Particularly important is the exchange of data from EGIB, GESUT, and BDOT500 with the databases of topographic objects (BDOT10k) and general geographic objects (BDOO). These databases contain information on the spatial location of objects and their descriptive attributes corresponding in detail to a topographic map at a scale of 1:10 000 and smaller scales. The current BDOT10k and BDOO databases are created and updated in accordance with the provisions of the Regulation on the database of topographic objects as well as the database of general geographic objects and standard cartographic studies of 2021.

The study of the degree of harmonisation of the EGIB, GESUT, and BDOT500 datasets with the BDOT10k database with regard to data quality and supply was performed by analysing the conceptual models of the spatial data registers under discussion. The study of the level of harmonisation of the datasets was carried out by classifying the analysed object classes into one of three groups, characterising the conformity of attribute values:

1. Full – high conformity in terms of spatial and descriptive definition and representation. Retrieval and provision of information (in principle) is automatic.
2. Partial – it is possible to acquire the geometry of objects and some descriptive information. Additional agreed operations are required.
3. None – there is no automatic data acquisition and supply.

3. Results and discussion

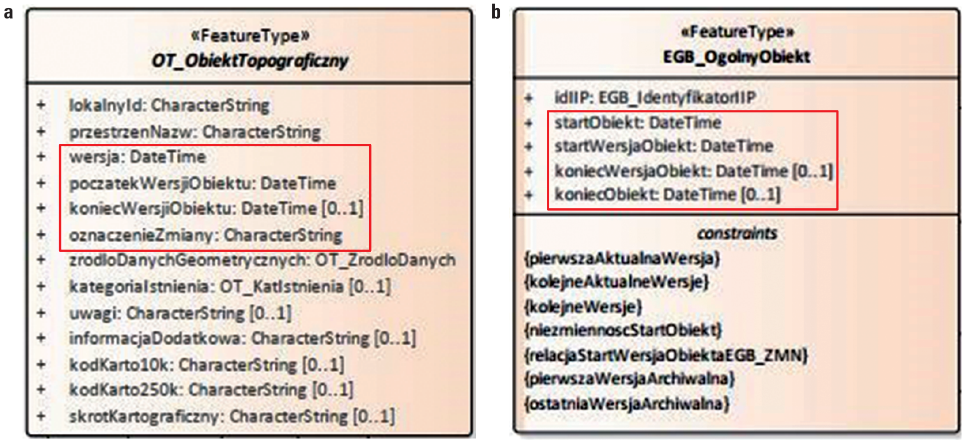
When analysing the degree of harmonisation of the conceptual models in terms of spatial data quality, the objects of the 'building' class (EGiB database) and the objects supplied by the GESUT and BDOT500 databases were studied. The conceptual models of spatial data quality for the 'building' class objects in the BDOT10k and EGiB databases show significant differences. The possibility of feeding quality properties from the EGiB database into BDOT10k is limited. BDOT10k clearly has a richer set of data quality features. For example, in EGiB the object 'building' does not have a property that directly describes the accuracy of the location, and there is only a reference to the technical operation instead (Table 1). Full agreement was found only for the date of object creation and the names of the data sources (geometric and attribute) of the origin of the object.

Table 1. Data quality properties of objects of the 'building' class in the BDOT10k database fed from EGiB

<table>
<tr><th colspan="2">Database: BDOT10k
Object class: building (OT_BUBD_A)</th><th colspan="2">Database: EGiB
Object class: EGiB_building</th><th rowspan="2">Value conformity</th></tr>
<tr><th>Property – name</th><th>Property – description</th><th>Property – name</th><th>Property – description</th></tr>
<tr><td>Date of creation</td><td>Date of creation of object</td><td>Start of life cycle of object</td><td>Date of creation of object</td><td>Full</td></tr>
<tr><td>Beginning of object's version</td><td>Date and time of the start of life cycle of given version of object</td><td>Start of life cycle of object's version</td><td>Date of creation of object's version</td><td rowspan="3">Partial</td></tr>
<tr><td>End of object's version</td><td>Date and time of the end of life cycle of given version of object</td><td>End of life cycle of object's version</td><td>Date of export of object's version to archive</td></tr>
<tr><td>–</td><td>–</td><td>End of life cycle of object</td><td>Date of export of object to archive</td></tr>
<tr><td>Geometric accuracy</td><td>Accuracy ≥ 1,5 m. For objects with difficult identification accuracy ≥ 5 m</td><td>–</td><td>Reference to the technical survey</td><td>Partial</td></tr>
<tr><td>Source of geometric data</td><td>EGiB, ortophotomap, base map, topographic map 10k, field survey and other</td><td>EGiB</td><td>EGiB database</td><td>Full</td></tr>
<tr><td>Source of attribute data</td><td>EGiB, ortophotomap, base map, topographic map 10k, field survey and other</td><td>EGiB</td><td>EGiB database</td><td>Full</td></tr>
</table>

Figure 1 shows excerpts from the UML application schemas for the BDOT10k and EGiB databases, indicating the differences in the attributes that define the life cycle of an object.

When analysing the conceptual models of spatial data quality for objects derived from GESUT and BDOT500 in relation to BDOT10k, significant inconsistencies were found. The supply of quality properties from the GESUT and BDOT500 databases is limited. For example, properties concerning position accuracy can be used, but only after additional alignment operations have been performed. Such supplementary operations include the determination of the interrelationship between geometric accuracy (BDOT10k) and the source of the object position data (GESUT and BDOT500) (Table 2). Full alignment was observed only for the date of the object and the names of the data sources (geometric and attribute) of the origin of the objects.



Source: Author's own study

Fig. 1. Diagrams of application of object lifetime registration for BDOT10k (Fig. 1a) and EGiB (Fig. 1b)

Table 2. Data quality properties of the BDOT10k database objects fed from the GESUT and BDOT500 databases

Database: BDOT10k Class: Topographic object (OT_ObiektTopograficzny)		Databases: GESUT, BDOT500 Classes: Object GESUT, object BDOT500 (GES_ ObiektGESUT), (OT_ ObiektBDOT500)		Value conformity
Property – name	**Property – description**	**Property – name**	**Property – description**	
Date of creation	Date of creation of object	Start of life cycle of object	Date of creation of object	Full
Beginning of object's version	Date and time of the start of life cycle of given version of object	Start of life cycle of object's version	Date of creation of object's version	Partial
End of object's version	Date and time of the end of life cycle of given version of object	End of life cycle of object's version	Date of export of object's version to archive	
–	–	End of life cycle of the object	Date of export of object to archive	

Table 2. cont.

Database: BDOT10k Class: Topographic object (OT_ObiektTopograficzny)		Databases: GESUT, BDOT500 Classes: Object GESUT, object BDOT500 (GES_ ObiektGESUT), (OT_ ObiektBDOT500)		Value conformity
Property – name	**Property – description**	**Property – name**	**Property – description**	
Geometric accuracy	Accuracy ≥ 1,5 m. For objects with difficult identification accuracy ≥ 5 m	Source of data on object's position	Measurement with a control network, measurement with a wire detector, map digitisation, et al.	Partial
Source of geometric data	GESUT, BDOT500, ortophotomap, base map, topographic map 10k, field survey and other	GESUT, BDOT500	GESUT or BDOT500 database	Full
Source of attribute data	GESUT, BDOT500, ortophotomap, base map, topographic map 10k, field survey and other	GESUT, BDOT500		Full

When analysing the degree of harmonisation of the conceptual models with regard to feeding the BDOT10k database with data from EGiB and GESUT, two classes of objects were studied: buildings and utilities networks. The EGiB database is the primary source of information on the spatial position of buildings and their attributes for the BDOT10k database. The required position accuracy for buildings in BDOT10k – 1.5 m – is ensured by practically acquired data, from any EGiB database (except for the cadastral map in the scale of 1:5000). When studying the conceptual models of the possibility of obtaining data for the BDOT10k database with the data on buildings from the EGiB, significant inconsistencies were found. The data supply from the EGiB database is limited. The EGiB database does not collect information on the detailed functions of buildings as defined in the Register of Classification of Measures (KŚT) and there is no 'building's proper name' attribute.

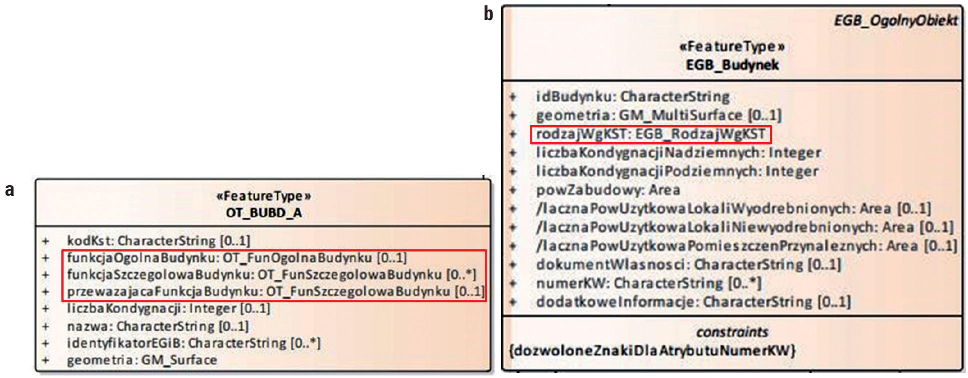
The extraction of building contours does not pose technical problems. Potential difficulties will arise, among other things, from topological errors of buildings with complex geometry [Lupa et al. 2018]. Performing additional operations to align attribute names will allow the extraction of information on the values of some attributes, such as the general and detailed function of the building and the number of storeys (Table 3).

Figure 2 presents excerpts from the UML application schemas for the BDOT10k and EGiB databases indicating the differences in the attributes defining building functions.

Table 3. Attributes of the objects of the 'building' class of the BDOT10k database supplied by the EGiB database

Databas: BDOT10k Object class: Budynek (OT_BUBD_A)		Database: EGiB Object class: EGiB_Budynek		Conformity of values of attributes
Attribute – name	**Attribute – description**	**Attribute – name**	**Attribute – description**	
Identificator	Acc to the idIIP	Identification	Acc to the idIIP	Full
Definition	Based on construction law	Definition	Based on construction law	Full
KŚT code	Code acc to the KŚT	Building type	Building type acc to the KŚT	Full
General function of building	General function of building acc to the KŚT	Building type	Building type acc to the KŚT	Full
Predominant detailed function of the building	Predominant detailed function of the building according to the KŚT	–	–	None
Detailed function of building	Detailed funtions of building according to the KŚT	–	–	None
Name	Building's proper name	–	–	None
Number of storeys	Number of above-ground storeys of building	Number of above-ground storeys	No. of the highest above-ground storey of building	None
EGiB identificator	Building identificatory acc to the EGiB database	EGiB	EGiB database	Full
Existance category	Object condition: operated/in construction/ destroyed/out of service	–	–	None
Geometry	Surface	Geometry	Polygon with enclaves	Partial

The GESUT database is the primary source of information on the spatial position of the utilities networks and their attributes for the BDOT10k database. The required position accuracy for the utility networks in BDOT10k – 1.5 m – is ensured by data obtained, in principle, from each GESUT database. With regard to the possibility of supplying the BDOT10k database with data on utilities networks from the GESUT, representative object classes were analysed: 'power line' and 'pipeline'. By examining the conceptual models, some inconsistencies were found. The data supply from the GESUT database is not fully automated. Both registers use consistent identifiers for objects, but it was allowed for GESUT to use local identifiers that cannot be used by BDOT10k (Table 4, Table 5).



Source: Author's own study

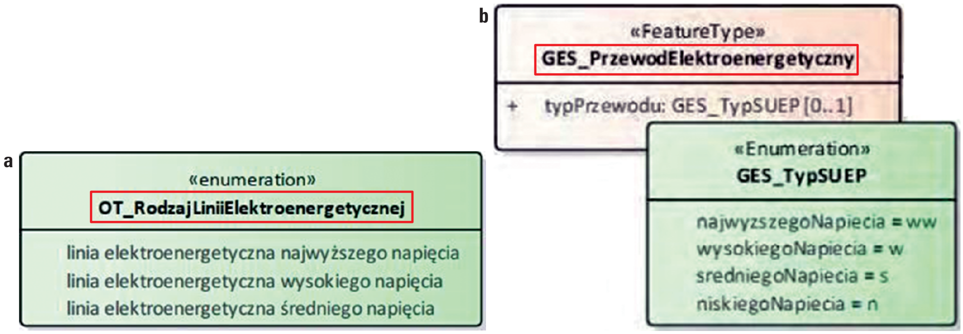
Fig. 2. Application diagrams of building function registration for BDOT10k (Fig. 2a) and EGiB (Fig. 2b)

The extraction of geometric axes of utilities networks does not present major technical problems. Potential difficulties will arise, among others, from topological and geometrical errors, such as the close proximity of vertices or different segmentation [Barańska et al. 2021]. Obtaining information on the values of some attributes, such as type, position, and diameter, requires additional alignment operations (Table 4, Table 5).

Table 4. Attributes of objects of the class 'power line' of the BDOT10k database supplied by the GESUT database

Database: BDOT10k Obejct class: power line (OT_SULN_L)		Database: GESUT Object class: power line (GES_PrzewodElektroenergetyczny)		Conformity of values of attributes
Attribute – name	**Attribute – description**	**Attribute – name**	**Attribute – description**	
Identificator	Acc to the idIIP	Identificator	Acc to the idIIP	Full
Definition	Based on construction law	Definition	Based on construction law	Full
Type	Power line: highest voltage/high voltage/ medium voltage	Type	Power line: highest voltage/high voltage/ medium voltage	Partial
Position	Above-ground	Course	Above-ground	Partial
Geometry	Line	Geometry	Line	Partial

Figure 3 presents excerpts from the UML application diagrams for the BDOT10k and GESUT databases with an indication of the differences in the attribute defining the power line.



Source: Author's own study

Fig. 3. Application diagrams of power line registration for BDOT10k (Fig. 3a) and GESUT (Fig. 3b)

Table 5. Attributes of objects of the class 'pipeline' of the BDOT10k database supplied by the GESUT database

Database: BDOT10k Object class: pipeline (OT_SUPR_L)		Databas: GESUT Object class: utilities cable (GES_Przewod)		Conformity of values of attributes
Attribute – name	**Attribute – description**	**Attribute – name**	**Attribute – description**	
Identificator	Acc to the idIIP	Identificator	Acc to the idIIP	Full
Definition	Based on construction law	Definition	Based on construction law	Full
Type	Petrol Heating Gas Sewage Oil Waterworks Other	Type	Petrol Heating Gas Sewage Oil Waterworks Other	Partial
Position	Ground Above-ground	Pipeline	Ground Above-ground	Partial
Diameter	Outer diameter of the pipeline	Diameter	Inner diameter of the pipeline	Partial
Geometry	Line	Geometry	Line	Partial

4. Conclusions

Making sure that the conceptual models of the official spatial databases are in mutual alignment is a real problem, and the current level of harmonisation impedes the effective exchange of information between the different registers. The EGiB, GESUT, and BDOT500 databases are characterised, with respect to BDOT10k, by a low level of harmonisation of conceptual models. This study has concluded that the retrieval and supply of information in the processes of creating and updating BDOT10k is limited. The extraction of information on the spatial location of objects does not present major technical problems. Inconsistencies in the conceptual models of the analysed registers concern mainly descriptive attributes. The BDOT10k database is characterised by a larger number of object classes and has richer attribute characteristics.

The correction of the conceptual model of the EGiB, GESUT, and BDOT500 databases in order to achieve a high degree of harmonisation with the BDOT10k model is feasible. All that is needed are appropriate institutional formal and legal actions.

Financed by a subsidy from the Ministry of Education and Science for the University of Agriculture in Krakow for 2023.

References

Baranowski M. 2012. Infrastruktura informacji przestrzennej w ujęciu systemowym. Seria monograficzna nr 15. Wyd. IGiK, Warszawa.

Barańska A., Bac-Bronowicz J., Dejniak D., Lewiński S., Krawczyk A., Chrobak T. 2021. A Unified Methodology for the Generalisation of the Geometry of Features. ISPRS, Int. J. Geo-Inf., 10, 107.

Directive 2007. Dyrektywa Parlamentu Europejskiego i Rady ustanawiająca infrastrukturę informacji przestrzennej we Wspólnocie Europejskiej (INSPIRE) z dnia 14 marca 2007 r. Nr 2007/2/WE.

Geodetic and Cartographic Law [2023] – Ustawa Prawo geodezyjne i kartograficzne z dnia 17 maja 1989 r. (Dz. U. z 2023 r., poz. 1752).

Kronenfeld B., Buttenfield B., Stanislawski L. 2020. Map generalisation for the future. ISPRS, Int. J. Geo. Inf., 9, 468.

Land and Building Register Ordinance of 2021 – Rozporządzenie Ministra Rozwoju, Pracy i Technologii z dnia 27 lipca 2021 r. w sprawie ewidencji gruntów i budynków (Dz. U. z 2023 r., poz. 1390).

Lupa M., Sarlej W., Adamek K. 2018. Harmonization of Datasets in the Frame of Spatial Data Infrastructure Using ETL Tools: A Case Study of BDOT500 and BDOT10k Databases. 2018 Baltic Geodetic Congress (BGC Geomatics), Olsztyn, Poland, 217–220.

Regulation on the geodetic register of utilities network of 2021 – Rozporządzenie Ministra Rozwoju, Pracy i Technologii z dnia 23 lipca 2021 r. w sprawie geodezyjnej ewidencji sieci uzbrojenia terenu (Dz. U. z 2023 r., poz. 1374).

Regulation on the database of topographic objects and the base map of 2021 – Rozporządzenia Ministra Rozwoju Pracy i Technologii z dnia 23 lipca 2021 r. w sprawie bazy danych obiektów topograficznych oraz mapy zasadniczej (Dz. U. z 2023 r., poz. 1385).

Regulation on the database of topographic objects as well as the database of general geographic objects and standard cartographic studies of 2021 – Rozporządzenia Ministra Rozwoju, Pracy i Technologii z dnia 27 lipca 2021 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów ogólnogeograficznych, a także standardowych opracowań kartograficznych (Dz. U. z 2023 r., poz. 1412).

Dr. Eng. Marek Ślusarski
Department of Land Surveying
Faculty of Environmental Engineering and Land Surveying
University of Agriculture in Krakow
ul. Balicka 253 a, 30-198 Kraków
e-mail: marek.slusarski@urk.edu.pl
ORCID: 0000-0002-8573-936X