

OpenStreetMap land cover data quality assessment on the example of Lower Silesia Voivodship, Poland

Jakub Łuczak^{1,*}

¹ Wrocław University of Science and Technology, Faculty of Geoengineering, Mining and Geology, 27 Wyb. Wyspiańskiego St., 50-370 Wrocław, Poland

Abstract. OpenStreetMap (OSM) is an open source, freely available spatial database, co-created by users from around the world in the idea of *volunteered geographic information*. The functioning of the project as an open community geographic information system is its great advantage, however, it is associated with many flaws, like heterogeneity of collected data. The presented work focuses on the assessment of completeness and quality of land cover data. The reference data used in analysis were objects stored in the *Baza Danych Obiektów Topograficznych* (BDOT10k), which is an element of the Polish National Geodetic and Cartographic Resource. The analysis was carried out for the area of the Lower Silesia Voivodship. Despite the achievement of quite unsatisfactory results of the analysis, OpenStreetMap project has information potential and is useful in selected spatial analyses.

1 Introduction

One of the components of spatial information systems is land cover data. These are terrain elements located on the surface of the Earth, created as a result of the activities of forces of nature or completely created by human economic activity. Information on land cover is extremely important to understanding the relationship between human and the environment [1]. This type of data is widely used in research of various scientific disciplines - both in natural sciences (such as biology, geography, ecology), and non-natural sciences such as urban planning [2].

There are several examples of land cover databases, like: Urban Atlas, which is an initiative of the European Commission in cooperation with the European Space Agency (ESA) and the European Environmental Agency (EEA), focused only on urban areas [2], CLC (Corine Land Cover) prepared by the European Union under the "Coordination of information on the environment" project covering the entire European Union with a maximum spatial resolution of 20 meters for selected 5 high-resolution layers [3] or NLCD (National Landcover Data) covering the continental part of the United States at a resolution of 30 meters. These collections are made available for free and widely used in many analyses. Land cover data are also part of state geodetic and cartographic resources, usually collected as vector data with accuracy higher than NLCD or CLC. An example of an official land cover dataset for the territory of Poland is the Topographic Data Database BDOT10k, whose elements were used as reference data in the analysis.

A kind of alternative to these databases are spatial data projects created by the users themselves. There is a wide range of different terminology being used to describe the creation of geospatial user-created content. e.g.: *crowdsourcing*, *collaboratively contributed geographic information*, *web-based public participation geographic information system*, *web mapping 2.0*, *neogeography* and *volunteered geographic information* (VGI) [4]. One of best examples of this type of project is OpenStreetMap. Data and maps created by OpenStreetMap users are published on the Open Database License (ODbL).

At least, one example of using OpenStreetMap data for creating a global land coverage database is known. This is the Open Land Cover (OLC) project, available at osmlanduse.com [5]. Data gaps in OLC were filled with free remote sensing data, but only for selected areas [6]. This paper, however, was based only on vector data coming directly from the OpenStreetMap database.

2 Material and methods

The aim of the article is to assess the geometric integrity of data collected in the OpenStreetMap database relating to land cover data from BDOT10k database. Only the geometry of the analyzed objects was analyzed, and the integrity of the attributes was not analyzed.

Previous works have usually focused on the analysis of network data [7, 8] or individual types of infrastructure, such as buildings [9, 10]. Examples regarding land cover were characterized by relatively small test areas [11]. Given this fact, the presented

* Corresponding author: jakub.luczak@pwr.edu.pl

results are innovative in the context of both the subject and the area of analysis.

Vector data analysis tools available in Esri's ArcMap environment were used during the study.

2.1 Reference data

Baza Danych Obiektów Topograficznych BDOT10k is a spatial database with details corresponding to a 1:10 000 topographic map, based on technical guidelines included in the *Regulation of the Minister of Interior and Administration of 17 November 2011 on the topographic objects database, the general geographic database and standard cartographic purposes*. The database collects information about topographic objects including:

- spatial location of objects in the national spatial reference system,
- objects characteristics,
- cartographical codes,
- metadata.

To ensure correct data exchange between different systems, object classification was made on three levels of detail [12]:

- Level 1: object class categories (9):
 - hydrographic network (code: *SW*),
 - communication network (*SK*),
 - utility infrastructure (*SU*),
 - land cover (*PT*),
 - building and constructions (*BU*),
 - landuse complexes (*KU*),
 - protected areas (*TC*),
 - administrative territorial division (*AD*),
 - other objects (*OI*),
- Level 2: object classes (from two to twelve in the category),
- Level 3: objects (from one to twenty-one in the class).

In the process of creating the BDOT10k database, public records are used (as a part of the Polish National Geodetic and Cartographic Resource). Records of other institutions and offices are also used in case of their usefulness and field inspections too (as a tool for supplementing and verifying data from those records).

In the study BDOT10k database valid for November 2013 was used. The database contractor determined the accuracy of the location of the land cover objects at 1.5 meters.

2.2 OpenStreetMap data

OpenStreetMap (OSM) is probably the most popular VGI project on the Internet [13]. As of November 8, 2017, the database had 4.1 billion nodes. The community creates over 4.3 million users, but only about 1 million made at least one edition (as of March 2018) [14].

OpenStreetMap data is organized in the database as a logical structure using XML. Within it, three basic elements can be distinguished: nodes, ways and relations. Each of these elements can be described by attributes, consisting of a key-value pair. The key and value can take any content, but the OpenStreetMap community has developed standards regarding the

characteristics of the most frequently mapped objects. The most important community-standardized classes of objects are [15]:

- communication ways: roads, railways, cycling ways and others (keys: *highway*, *railway*, *cycleway*, *aerialway*),
- natural watercourses and water objects made by man (*waterway*),
- landuse and natural objects (*landuse*, *natural*, *geological*),
- buildings, barriers (*man_made*, *energy*),
- facilities related to the generation and distribution of energy (*power*),
- public transport facilities (*public_transport*),
- recreational and sports facilities, amenities (*tourism*, *leisure*, *amenity*),
- public facilities, production sites and services (*emergency*, *office*, *shop*, *history*, *craft*),
- administrative boundaries (*boundary*),
- routes as non-physical objects (*route*).

Due to the widespread use of OpenStreetMap database elements as a basemaps in social projects or commercial solutions, separate classes for public facilities, recreational facilities or obstacles affecting various spheres of everyday life, usually not included in official reports, are used. It should be noted that due to the dynamics of changes of objects mapped in OpenStreetMap project, the suggested structure of attributes of these classes changes most often.

Data collected in the OpenStreetMap database is characterized by heterogeneous geometric accuracy, depending on various methods and techniques of their acquisition [9]. The mean error of a position of objects is between 3-8 meters with maximum deviations of 20 meters for the analysis of the accuracy of the London road network in compared to Ordnance Survey database [7].

In the study OpenStreetMap database valid for May 2018 was used.

2.3 Study area

The analysis was carried out for the area of the Lower Silesia Voivodship – it is area of 19,947 sq. km (see Figure 1). It is characterized by a varied landscape structure, including each of the aforementioned categories of object classes. For this reason, this area is a good example of the possibility of applying the proposed methodology and referring the results to the rest of the country.

2.4 Data comparison

"Land cover" is defined as the most important surface situational elements of the terrain, distinguishable on the basis of their external view (physiognomic features), and not their functions [16]. Objects belonging to this category maintain a neighborhood relation to each other and in the BDOT10k database they describe the whole area in a continuous and complete way [16]. In turn, data stored in the OpenStreetMap database regarding land

cover elements usually do not have continuity, additionally, the situation of overlapping objects of different classes may occur.

All 12 land cover classes were used in analysis (dataset was continuous in the entire study area). Each

class has been assigned to commonly used and standardized by OpenStreetMap community keys and values [5, 15]. Comparison of classes from reference database and from OpenStreetMap is shown in Table 1.

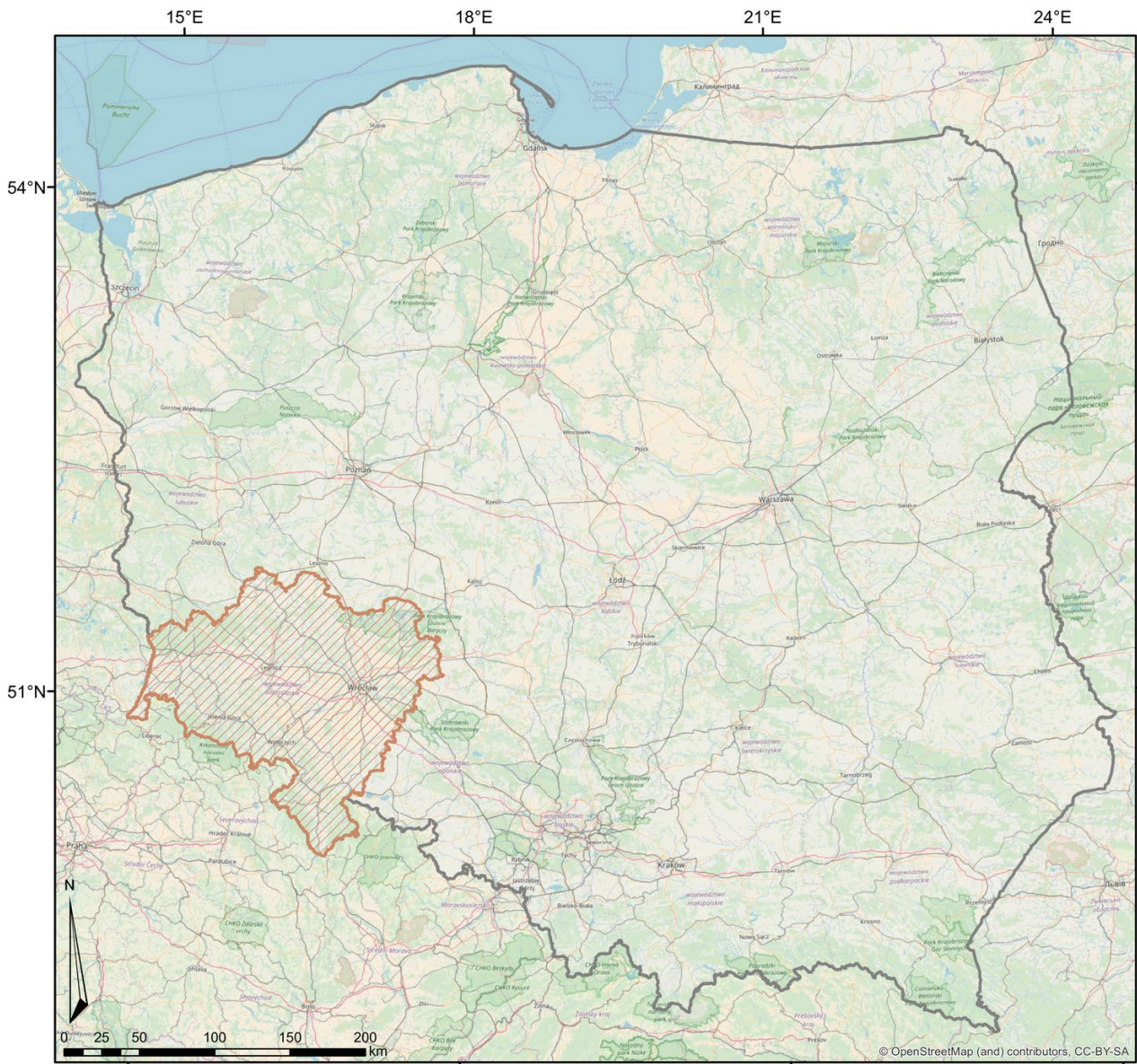


Fig. 1. Study area (marked) compared with Poland borders with OpenStreetMap data as basemap.

Table 1. Object classes compared in the study.

object class category (BDOT10k code)	object class (BDOT10k code)	objects (in BDOT10k)	OpenStreetMap key and value
land cover (PT)	surface water (PTWP)	seas flowing water standing water	landuse=basin landuse=reservoir natural=water waterway=riverbank
	development (PTZB)	multi-family residential single-family residential industrial development commenrcial development	landuse=commercial landuse=construction landuse=farmyard landuse=garages landuse=industrial landuse=residental landuse=retail
	woodland (PTLZ)	forest wood grove	landuse=forest natural=wood
	shrubland (PTRK)	scrub	natural=scrub
	pernament crops (PTUT)	allotments plantation orchard forest nurseries plant nurseries	landuse=allotments landuse=greenhouse_horticulture landuse=orchard landuse=plant_nursery landuse=vineyard
	grasses and arable farmland (PTTR)	grasses arable farmland	landuse=farmland landuse=grass landuse=meadow natural=grassland natural=heath
	terrains under roads, railways and airports (PTKM)	terrains under roads terrains under railways terrains under airways	aeroway=aerodrome aeroway=apron aeroway=heliport aeroway=runway landuse=railway
	unutilized land (PTGN)	scree, mound or rubble rocky land sandy land other unused land	landuse=brownfield natural=bare_rock natural=sand natural=scree natural=shingle
	yards, sites (PTPL)	yards, sites	<i>none</i>
	landfill (PTSO)	municipal landfill industrial landfill	landuse=landfill
	working, excavation heap (PTWZ)	working excavation heap	landuse=quarry
	other undeveloped areas (PTNZ)	terrain under constructions and other buildings cemeteries other areas	landuse=cemetery landuse=greenfield

3 Results

In first step the corresponding data were compared only in terms of area in proportion to whole study area. Coverage with OpenStreetMap data used in analysis amounted to 57% of study area (see Figure 2). The results which are summarized in Table 2, show that only a few classes have reached a data ratio close to 1. These are classes related to the natural coverage (surface waters, forest and wooded areas) and areas relatively clearly identifiable on aerial images or in situ measurements (working, excavation heap). The value of the ratio for the other classes reaches either very high values (>3) or exceptionally low values (<0.5). This is connected with semantic flaw of individual classes - it is

possible to indicate groups of classes which, due to physical similarity of mapped areas, are classified by editors differently than reference data (eg PTPL-PTZB, PTRK-PTLZ, PTNZ-PTZB, see Table 3). The possible reason for this situation is too detailed specification of anthropogenic facilities. Defining classes in the reference database should be connected to the most important surface elements, distinguishable on the basis of physiognomic features, and not functions performed [17]. This is also confirmed by the results achieved for easy identifiable natural data such as water or wooded areas, for which the highest consistency and the smallest data gaps were obtained. Differences in the timeliness of the data did not have a significant impact during

comparisons (land cover is a relatively stable element of natural environment).

It is worth emphasizing that two classes of objects (terrains under roads, railways and airports, permanent crops) are characterized by the lack of OpenStreetMap data exceeding 60%. In the case of the first of them, the reason is the way of transport network data collection in OpenStreetMap database (practically only as linear

objects), and in the case of crops - this layer is kind of „background” layer (most unmapped contents are crops). In the case of other classes, the gaps in the data amount to 20-40%, so they are significant values too.

Table 3 presents relationships between OpenStreetMap data and reference data. Values for directly corresponding classes are marked.

Table 2. Comparison of BDOT10k and OpenStreetMap object classes.

object class	area in BDOT10k dataset, ha	area of corresponding data in OSM dataset, ha	ratio
water bodies	24 873.1	27 121.4	1.09
development	70 399.6	110 377.3	1.57
woodland	665 778.6	619 887.1	0.93
shrubland	3 411.4	12 404.6	3.64
pernament crops	18 627.3	9 711.4	0.52
grassess and arable farmland	1 177 837.5	354 244.5	0.30
terrains under roads, railways and airports	17 134.4	734.4	0.04
unutilized land	3 224.1	473.6	0.15
yards, sites	3 341.9	0.0*	0.00
landfill	393.1	1 641.1	4.18
working, excavation heap	5 657.9	6 361.0	1.12
other undeveloped areas	4 024.2	995.9	0.25
no data	0.0**	850 747.9	N/A
Σ	1 994 703.2	1 994 703.2	1.00

* - no corresponding data keys in OpenStreetMap, ** - BDOT10k is continuous dataset

Table 3. Contingency table for BDOT10k and OpenStreetMap data.

		corresponding OpenStreetMap data (see Table 1), %												
		PTWP	PTZB	PTLZ	PTRK	PTUT	PTTR	PTKM	PTGN	PTPL	PTSO	PTWZ	PTNZ	none
BDOT10k data	PTWP	99.584	0.013	0.214	0.012	0.054	0.023	0.002	0.000	0.000	0.001	0.001	0.032	0.064
	PTZB	0.022	75.108	0.605	0.082	0.136	1.544	0.071	0.023	0.000	0.008	0.085	0.034	22.282
	PTLZ	0.057	0.510	88.352	0.632	0.021	0.417	0.002	0.011	0.000	0.122	0.073	0.016	9.787
	PTRK	0.389	5.248	20.105	32.196	0.199	5.163	0.002	0.072	0.000	0.056	0.516	0.002	36.052
	PTUT	0.050	16.612	1.984	0.865	45.505	6.504	0.010	0.008	0.000	0.000	0.023	0.022	28.417
	PTTR	0.111	3.769	2.415	0.563	0.077	29.631	0.030	0.013	0.000	0.033	0.011	0.008	63.339
	PTKM	0.021	17.091	5.114	0.617	0.099	4.033	1.562	0.007	0.000	0.000	0.112	0.277	71.067
	PTGN	5.878	2.878	30.272	2.584	0.013	4.133	0.174	1.816	0.000	0.166	10.043	0.227	41.816
	PTPL	0.046	54.204	0.872	0.230	0.168	2.548	0.494	0.376	0.000	0.090	0.678	0.000	40.294
	PTSO	0.143	13.844	1.553	2.937	0.279	5.686	0.000	0.799	0.000	45.332	6.190	0.342	22.895
	PTWZ	2.398	3.812	2.173	0.561	0.008	2.308	0.000	2.477	0.000	5.519	68.050	0.001	12.693
	PTNZ	5.029	34.697	0.948	0.372	1.556	2.265	0.667	0.301	0.000	0.650	17.467	13.806	22.242

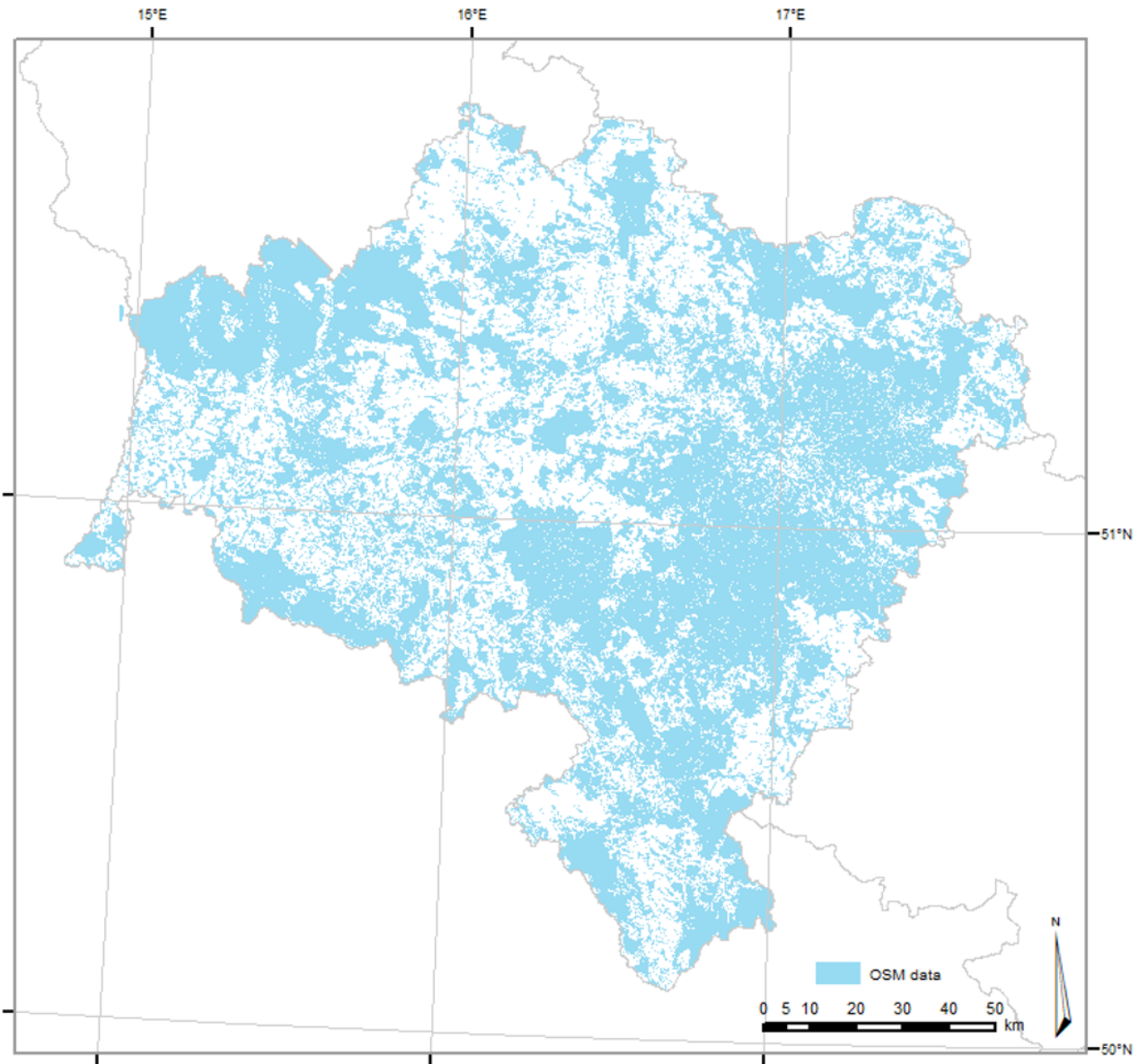


Fig. 2. Cover the study area by OpenStreetMap data used in the study.

4 Conclusions

The results obtained during analysis allow to conclude that data from the OpenStreetMap database referring to the territory of Poland (BDOT10k database) are still not valuable information about the land cover data. Due to the values obtained in the comparison of individual classes (for the 8 contingency classes values below 50%) and coverage at 57% level, OpenStreetMap data cannot be used as an alternative to the BDOT10k database, which is the basis for 1:10 000 and smaller maps. The reasons for these unsatisfactory results are: the differences in the classification of land cover classes, the heterogeneous level of data detail and the degree of incompleteness of data in the OpenStreetMap and BDOT10k databases. However, it seems that it is possible to use individual object classes for some simple land cover analyzes on a macro scale and in creating thematic maps. To sum up: despite the quality

inconveniences, the OpenStreetMap database has huge, constantly growing information potential and in the future, in the case of specifying and normalizing the standards of data about land cover collection, it may be really valuable data source.

References

1. M. Borowska-Stefańska, K. Leśniewska-Napierała, S. Wiśniewski, *Geografie*, **123**, 63-83 (2018)
2. A. Dąbrowski, *Metodyka opracowywania szczegółowych map pokrycia terenu na podstawie istniejących źródeł danych*. in: M. Nowak (ed.): *GIS i dane przestrzenne w ocenach oddziaływania na środowisko. Podręcznik dobrych praktyk* (Wydawnictwo Naukowe, Poznań, 2016)
3. Chief Inspectorate of Environmental Protection, *CORINE Land Cover – CLC 2012*.

- <http://clc.gios.gov.pl/index.php/clc-2012/o-projekcie>, Retrieved May 30, 2018
4. S. Fritz, I. McCallum, C. Schill, R. Grillmayer, F. Achard, F. Kraxner, M. Obersteiner. *Remote Sensing*, **1**, 3, 345-354 (2009)
 5. OSM Lalduse Landcover, *Info&Contact*. <http://osmlanduse.org/#12/17/51/0/>, Retrieved May 27, 2018.
 6. M. Schultz, J. Voss, M. Auer, S. Carter, A. Zipf, *International Journal of Applied Earth Observation and Geoinformation*, **63**, 206-213 (2017)
 7. M. Haklay. *Environment and Planning B: Planning and Design*, **37**, 682-703 (2010)
 8. D. Zielstra, H. Hochmair, P. Neis. *Transactions in GIS*, **17**, 315-334 (2013)
 9. J. Nowak Da Costa, E. Bielecka, B. Całka. *Roczniki Geomatyki*, **14**, 201-211 (2016)
 10. Z. Qi, *International Journal of Geographical Information Science*, **32**, 257-281 (2018)
 11. C.C. Fonte and B. Martinho, *International Journal of Geographical Information Science*, **31**, 2382-2400 (2017)
 12. Ministry of Interior and Administration, *Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 17 listopada 2011 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów ogólnogeograficznych, a także standardowych opracowań kartograficznych* (2011)
 13. M.A. Brovelli, M. Minghini, M. Molinari, P. Mooney, *Transactions in GIS*, **21**, 191-206 (2017)
 14. OpenStreetMap Wiki, *OSM Stats*. <https://wiki.openstreetmap.org/wiki/Stats>, Retrieved May 11, 2018
 15. OpenStreetMap Wiki, *Map Features*. https://wiki.openstreetmap.org/wiki/Map_Features, Retrieved May 11, 2018
 16. D. Gotlib, *Ogólna koncepcja, cel budowy i zakres informacyjny BDOT10k i BDOO*. in: *Rola bazy danych obiektów topograficznych w tworzeniu infrastruktury informacji przestrzennej w Polsce*, Head Office of Land Surveying and Cartography, Warsaw, 51-83 (2013)
 17. J. Bac-Bronowicz, A. Głażewski, P. Liberadzki, I. Wilczyńska, *Roczniki Geomatyki*, **13**, 295-305 (2015)