

ELŻBIETA BIELECKA, MONIKA LESZCZYŃSKA

Użyteczność danych o lasach udostępnianych w serwisie OpenStreetMap*

Usability of the OpenStreetMap forest data

ABSTRACT

Bielecka E., Leszczyńska M. 2018. Użyteczność danych o lasach udostępnianych w serwisie OpenStreetMap. Sylwan 162 (6): 460-468.

OpenStreetMap (OSM) is a collaborative project covering the whole World with topographic data. Data are edited by volunteer users on the basis of satellite images, ortophotomaps and GPS traces, therefore are of diversified quality, especially completeness and positional accuracy. The aim of this study was to evaluate the usability of OSM forest data for Poland. We proved that positional accuracy, expressed by RMSE that equals to 16 m, was comparable with the national topographic data at the scale 1:10 000. However 3.5% of the outliers were noticed, for which the positional accuracy amounted to about 300 m. The completeness of the OSM data in May 2017, computed on the basis of Czekanowski matching feature area-based completeness index, equalled 86% and was still increasing. The lowest value of the completeness was observed for lubuskie voivodeship (76.4%), while the highest (92.5%) for zachodniopomorskie. For ten voivodeships the completeness of forest cover was greater than national average. In general, there were only a few forest polygons with their name or other pertinent characteristics like tree type (coniferous/deciduous) attributed by the editors. Logical consistency was nearly perfect as only 5.8% of the polygons were topologically incorrect. Forest definition as 'an area dominated by trees', is very similar in national topographic data and in OSM, both are based on physical characteristics of the surface of the Earth. Contrary to the national spatial data, OSM data are constantly updated, free and available without restrictions. It constitutes valuable alternative data source for all users. However, the heterogeneity of these data should also be kept in mind, while they are used in analysis and decision process support.

KEY WORDS

forest, OSM, spatial data quality, completeness, location accuracy

ADDRESSES

Elżbieta Bielecka – e-mail: elzbieta.bielecka@wat.edu.pl
Monika Leszczyńska

Wydział Inżynierii Lądowej i Geodezji, Wojskowa Akademia Techniczna; ul. gen. W. Urbanowicza 2, 00-908 Warszawa

Wstęp

Lasy zajmują prawie $\frac{1}{3}$ powierzchni Polski. Informacja o ich przestrzennym rozmieszczeniu jest bardzo istotna dla administracji publicznej i sektora prywatnego oraz edukacji i nauki. Jest wykorzystywana w różnego rodzaju analizach przestrzennych, a także przedstawiana na mapach,

*Artykuł powstał w ramach prac statutowych nr PBS/333/2017 realizowanych na Wydziale Inżynierii Lądowej i Geodezji Wojskowej Akademii Technicznej w Warszawie.

w tym topograficznych, turystycznych czy tematycznych. Zapotrzebowanie na informacje o rozmieszczeniu i stanie lasów ma odzwierciedlenie w rosnącej liczbie zbiorów danych zawierających dane o lasach. Zbiory takie są tworzone i zarządzane nie tylko przez Lasy Państwowe, ale także przez służbę geodezyjną i kartograficzną, powstają również w ramach międzynarodowych projektów, m.in. Global Land Cover, MODIS Global Land Cover, CORINE Land Cover, LUCAS czy OpenStreetMap. Duża dostępność danych przestrzennych o lasach stawia użytkownika przed trudnym wyborem zbioru najlepiej spełniającego jego oczekiwania, a zatem przed koniecznością określenia jego użyteczności. Użyteczność danych przestrzennych została zdefiniowana w normie ISO 19157 Geographic information – Data quality, w której jest traktowana jako jeden z elementów jakości danych. Zgodnie z tym podejściem może być opisywana za pomocą: kompletności, spójności (logicznej, topologicznej, dziedziny i formatu), dokładności położenia, dokładności tematycznej, a także innych charakterystyk uznanych przez producenta danych za istotne. Z punktu widzenia użytkownika do najważniejszych charakterystyk zbiorów danych przestrzennych należą: zakres geograficzny i tematyczny danych, rozdzielczość przestrzenna, dokładność lokalizacji oraz dokładność tematyczna [Bielecka 2015]. Duże znaczenie mają także informacje mówiące o zasadach wykorzystania danych (rodzaj licencji) i ograniczeniach związanych z danymi (np. wysoka opłata, udostępnianie tylko części danych). Wszystkie te informacje można znaleźć w metadanych, o ile zbiory zostały nimi opisane.

Z analiz wykonanych przez Hościło i in. [2016] wynika, że użyteczność danych o lasach udostępnianych przez organy administracji publicznej (Lasy Państwowe, Główny Urząd Geodezji i Kartografii, Agencję Restrukturyzacji i Modernizacji Rolnictwa oraz Główny Urząd Statystyczny), mierzona dokładnością określenia powierzchni lasów, wykazuje różnice. Dochodzą one do 8 tys. km², co stanowi prawie 9% powierzchni lasów podawanej przez GUS. Różnice te wynikają z kilku powodów, w tym m.in. odmiennego definiowania lasu w różnych zasobach, rodzaju zasobu ze szczególnym uwzględnieniem dokładności gromadzonych danych oraz stabilności przepisów regulujących zakres i szczegółowość gromadzonych danych. Problem ten kompleksowo rozważa Jabłoński [2015a, b]. Podkreśla on, że definicje lasu według FAO, Forest Europe oraz konwencji klimatycznej z Kioto nie są zgodne z definicją krajową zawartą w ustawie o lasach. Na niespójność definicji lasu zwracają uwagę także Björk i Skånes [2015] oraz Luc i Bielecka [2015]. W wielu zbiorach danych zawierających informacje o lasach stosuje się definicję bazującą na klasyfikacji pokrycia terenu, zgodnie z którą łącznie przedstawia się las i obszary zalesione. Jest to szczególnie widoczne wszędzie tam, gdzie podstawą pozyskania informacji o lasach są zdjęcia lotnicze i obrazy satelitarne. Las jest wówczas definiowany jako obszar porośnięty drzewami o określonej powierzchni minimalnej (zależnej od rozdzielczości przestrzennej danych, np. w przypadku CORINE Land Cove 25 ha), stopniu pokrycia koronami (najczęściej minimum 30-50%) i wysokości drzew (z reguły wynoszącej więcej niż 2-5 m).

Różnice między powierzchnią lasów zawartą w ewidencji gruntów a stanem faktycznym wynikają głównie z wieloletnich zaniedbań w ustalaniu granic użytków i są usuwane w wyniku modernizacji ewidencji gruntów i budynków [Noszczyk, Hernik 2016; Izdebski 2017]. Niewątpliwie do ich powstania przyczyniła się także zmiana w 2012 roku klasyfikacji użytków gruntowych. Jabłoński [2015b], analizując zalesienia w ramach PROW, doszedł do wniosku, że różnice w powierzchni lasów wynikają także z przekwalifikowania terenów zalesionych na lasy dopiero w czwartym lub piątym roku po zalesieniu.

OpenStreetMap (OSM) jest ogólnosiwiatowym projektem społeczności internetowej, którego celem jest stworzenie darmowej oraz powszechnie dostępnej mapy całej kuli ziemskiej. Do jej rozwoju przyczyniła się technologia Web 2.0 umożliwiająca dodawanie przez użytkow-

ników własnych treści do serwisów internetowych. W bazie przechowywana jest historia edycji każdego obiektu, co pozwala na usuwanie pomyłek i aktów wandalizmu. Fenomen danych tworzonych przez użytkowników, określanych ogólnie jako VGI (Volunteered Geographical Information), w tym głównie danych OSM, jest od wielu lat przedmiotem badań. Prace badawcze skupiają się w dwu głównych nurtach: ocenie jakości danych i charakterystyce użytkowników – edytorów wprowadzających dane do serwisu OSM. Ocena jakości dotyczy z reguły kompletności i dokładności lokalizacji w odniesieniu do innych, bardziej wiarygodnych danych, przy czym najczęściej badanymi obiektami są budynki, drogi oraz punkty charakterystyczne (POI). Z analiz tych jednoznacznie wynika, że pokrycie danymi OSM jest bardzo zróżnicowane, większe na obszarach miast i znacznie mniejsze na terenach wiejskich [Girres, Touya 2010; Haklay 2010; Zielstra, Zipf 2010; Ludwig i in. 2011; Neis i in. 2013]. Wnioski te potwierdzają także analizy jakości danych OSM wykonane dla Polski [Cichociński 2012; Marczak 2015; Nowak Da Costa i in. 2015]. Charakterystyka edytorów dotyczy najczęściej ich zaangażowania w budowę OSM [Budhathoki 2010; Neis, Zipf 2012] oraz zależności pomiędzy lokalizacją edytora a kartowanym obszarem [Budhathoki 2010; Arsanjani, Bakillah 2015]. Pojawiają się także publikacje analizujące motywacje i zachowania twórców OSM [Schmidt, Klettner 2013; Steinmann i in. 2013]. Edytorzy OSM są grupą mocno zróżnicowaną zarówno ze względu na kierujące nimi motywacje, jak i geograficzną lokalizację obszarów, które aktualizują. Większość współtwórców OSM to ludzie młodzi, wykształceni, mieszkający w dużych miastach. Kartują oni najczęściej okolice miejsca zamieszkania, pracy lub czasowego przebywania, wykorzystując dane satelitarne i metodę wektoryzacji, rzadziej ślady GPS. Podkreśla to także Marczak [2017], badając m.in. profil współtwórców OSM z Polski. Główne zalety OSM to: udostępnianie danych zgodnie z licencją Open Database License (ODbL), ciągła aktualizacja oraz dostęp do danych dostarczanych w powszechnie używanych formatach shp lub xml. Sprawia to, że wprowadzone dane są aktualne oraz łatwo i szybko dostępne. Do wad należy duża heterogeniczność w zakresie kompletności wprowadzanych danych oraz ich geometrycznej dokładności. Jest to szczególnie uciążliwe dla obszarów rolniczych, wsi i małych miast.

Celem badań była ocena użyteczności serwisu OSM w zakresie danych o lasach. Zdefiniowano hipotezę badawczą mówiącą, że społecznościowe, ogólnodostępne dane o lasach dostarczane przez serwis OpenStreetMap są wiarygodnym źródłem danych, porównywalnych z danymi urzędowymi.

Material i metody

Oceniając użyteczność danych o lasach OpenStreetMap, skupiono się na określeniu ich: kompletności, spójności (pojęciowej, topologicznej, formatu), dokładności położenia oraz dostępności i pochodzenia. W badaniach wykorzystano metodę bezpośredniej oceny, polegającą na analizie danych zawartych w zbiorze (metoda wewnętrzna), jak też analizie porównawczej ze zbiorem uznanym za wiarygodniejszy (metoda zewnętrzna). Jako zbiory referencyjne przyjęto Bazę Danych Obiektów Topograficznych (BDOT10k), Bazę Danych Obiektów Ogólnogeograficznych (BDOO) oraz ortofotomapy dostępne w serwisie geoportal.gov.pl.

Kompletność i spójność oceniono dla wszystkich obiektów znajdujących się w bazie. Jest to kontrola pełna. Wyniki przedstawiono w procentach informujących o udziale poprawnie wprowadzonych do OSM obszarów (wieloboków) leśnych w stosunku do danych referencyjnych. Przy obliczeniach kompletności obiektów został zastosowany współczynnik kompletności odpowiadający sobie semantycznie zbiorów (*CC&I*), bazujący na współczynniku podobieństwa Czekanowskiego (ang. matching feature area-based completeness index), zastosowanym po raz

pierwszy przez Nowak Da Costę [2016]. Przewaga tego wskaźnika nad innymi dostępnymi w literaturze (np. współczynnikiem Jacarda) polega na porównaniu tylko i wyłącznie obiektów tożsamyh semantycznie. Dokładność położenia określono na podstawie równomiernie rozłożonej próby 200 obszarów leśnych. Jest to kontrola niepełna, statystyczna. Dla wybranych obiektów pomierzono wektory przesunięcia (Δx i Δy) granicy lasu danych OSM w stosunku do lasu na ortofotomapie oraz obliczono błąd położenia, wektory przesunięć Δd i współczynnik zmienności v pokazujący rozproszenie długości wektorów przesunięcia granicy w stosunku do długości średniej.

Dane OSM pozyskano ze strony <http://download.geofabrik.de/europe/poland.html> w maju 2017 roku. Lasy w OSM, definiowane jako obszar porośnięty drzewami, mogą być elementem warstw landuse i natural (<https://wiki.openstreetmap.org/wiki/Tag:landuse%3Dforest>). W kategorii landuse lasy są oznaczane tagiem (atrybutem) „forest”. Zakwalifikowane są tu przede wszystkim lasy wykorzystywane gospodarczo. W kategorii natural powinny się znajdować tylko lasy naturalne, niewykorzystywane gospodarczo i niepielęgnowane, oznaczane jako „wood”. Należą do nich lasy w granicach rezerwatów i parków narodowych. Administratorzy OSM ostrzegają przed stosowaniem tagu „landuse=wood”, uważając, że jest on semantycznie niepoprawny.

BDOT10k i BDOO to urzędowe dane topograficzne, których szczegółowość i dokładność lokalizacji odpowiada skalom 1:10 000 i 1:200 000. Aktualność BDOT10k to lata 2012-2014 (w zależności od regionu kraju), BDOO podobnie, ponieważ baza ta powstała w wyniku generalizacji danych BDOT10k. Lasy są częścią klasy „pokrycie terenu”, oznaczoną kodem PTLZ. Należą do niej tereny leśne i zadrzewione definiowane jako tereny o zwartym zadrzewieniu, czyli „lasy, zadrzewienia parków i cmentarzy oraz inne tereny zadrzewione, w szczególności przylegające do wód powierzchniowych grunty porośnięte drzewami, zadrzewienia terenów lotniskowych, rekreacyjnych” [Rozporządzenie... 2011]. W BDOT10k znajdują się lasy i obszary zadrzewione, jeśli ich powierzchnia wynosi co najmniej 0,1 ha, a minimalna szerokość i długość wynoszą odpowiednio 10 i 40 m. W BDOO znajdują się obszary leśne o szerokości co najmniej 125 m i powierzchni większej od 25 ha. Jednakże Rozporządzenie... [2011] dopuszcza wyjątki od tej reguły, w wyniku czego w obu zbiorach znajduje się wiele obszarów o powierzchni mniejszej niż zalecana. Dane BDOO i BDOT10k wykorzystano do analizy kompletności obiektów (wieloboków zaklasyfikowanych jako lasy) oraz atrybutów. Ortofotomapy o pikselu od 0,5 do 0,1 m są dostępne w serwisie geoportal.gov.pl. Większość ortofotomap przedstawia stan powierzchni terenu w latach 2015-2016, jedynie dla północno-zachodniej części Polski oraz fragmentu województwa mazowieckiego dane te pochodzą z lat 2010-2012. Ortofotomapy wykorzystano do określenia dokładności lokalizacji kompleksów leśnych. Wszystkie zbiory referencyjne są rejestrami publicznymi dostępnymi w państwowym zasobie geodezyjno-kartograficznym, jednocześnie są danymi Infrastruktury Informacji Przestrzennej w Polsce [Bielecka, Medyńska-Gulij 2015].

Wyniki

Powierzchnia lasów w Polsce według OSM wynosi 9284,98 tys. ha i jest większa od powierzchni podawanej przez GUS o 87 tys. ha (0,9% powierzchni lasów w kraju). Stanowi ona ponad 86% powierzchni lasów zarejestrowanych w BDOT10k. Jednocześnie jest mniejsza o prawie 713 tys. ha od powierzchni lasów określonej przez Hościło i in. [2016] na podstawie wieloźródłowej analizy danych i nazwanej rzeczywistą powierzchnią lasów.

Kompletność lasów w bazie OSM w poszczególnych województwach jest w niewielkim stopniu zróżnicowana, na co wskazuje współczynnik zmienności wynoszący 2,42%. Najwięcej

lasów wprowadzono do OSM na terenie województwa opolskiego ($CCzI=92,53\%$), na kolejnych miejscach są województwo zachodniopomorskie ($CCzI=89,72\%$) i wielkopolskie ($CCzI=89,42\%$). Aż w dziesięciu województwach kompletność jest większa od średniej dla Polski, wynoszącej 86,18%. Wartość kurtozy 2,65 świadczy o większej niż normalnie koncentracji wartości kompletności wokół średniej. Najmniej lasów, w stosunku do stanu faktycznego, wprowadzono na obszarze województwa lubuskiego ($CCzI=76,4\%$), kompletność poniżej średniej dotyczy sześciu województw (ryc.).

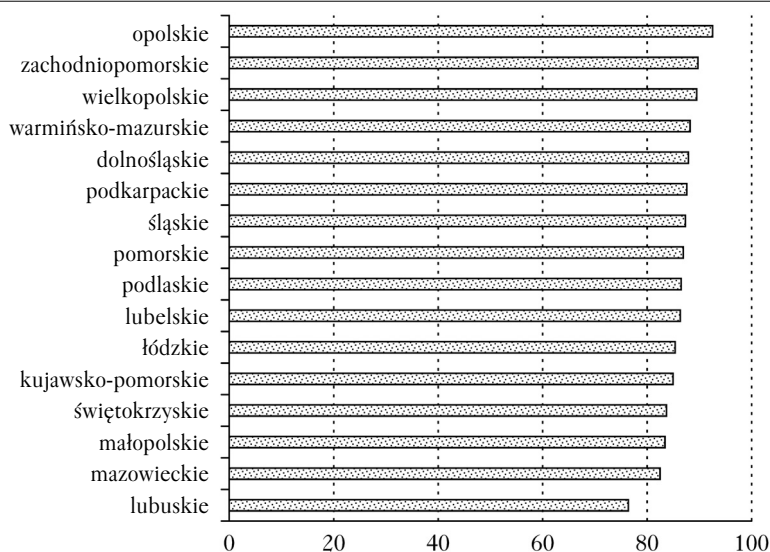
Zakres tematyczny OSM jest dość ubogi. Lasy możemy dodatkowo opisać nazwą, niestety czyni to niewielu twórców OSM. Całkowita liczba wieloboków zaklasyfikowanych jako lasy w bazie OSM wynosi 204 114, z czego zaledwie 1046 ma nazwę (0,5%), w tym 61 nazw podanych jest błędnie (5,8%). W bazach BDOT10k i BDOO atrybutów tematycznych jest znacznie więcej, informują one użytkownika m.in. o kategorii drzewostanu (np. mieszany), źródle danych, kategorii dokładności geometrycznej oraz nazwie, jeśli taka istnieje.

W danych OSM dla Polski wszystkie lasy znajdują się w warstwie landuse i mają przypisany atrybut „forest”. W warstwie natural nie ma obiektów z atrybutem „wood” ani „forest”, charakteryzują się więc one 100-procentową spójnością pojęciową. Ponadto semantycznie odpowiadają definicji lasów stosowanej w BDOT10k i BDOO.

Spójność topologiczna wynosi 96,4%. Wykryto 7393 wieloboki (5,5%) niespełniające reguły ich niepokrywania się. Stosunkowo najmniej błędów topologii było w lasach województwa podkarpackiego (zaledwie 113), najwięcej zaś w województwie podlaskim – 1023 wieloboki.

Dane OSM są dostępne na podstawie otwartej licencji ODbL w postaci xml, osm.pbf lub shp. Ponadto historia tworzenia danych jest zapisana w plikach osh.pbf. Wszystkie pliki są generowane automatycznie co kilkanaście godzin, w zależności od aktywności edytorów. W pracy wykorzystano pliki shp, wobec czego spójność formatu wynosi 100%.

Dokładność lokalizacji lasów w OSM jest zróżnicowana (tab. 1). Bardzo duży rozstęp pomiędzy długością wektorów przesunięć (Δx , Δy , Δd) świadczy o występowaniu błędów grubych.



Ryc.

Completność danych OSM w województwach (maj 2017)
Completness of the OSM data (May 2017)

Błędy takie zaobserwowano dla siedmiu wieloboków, co stanowi 3,5% badanej próby. Jest to istotne spostrzeżenie, potwierdzające wyniki analiz dokładnościowych raportowane w literaturze przedmiotu. Bardzo duże zróżnicowanie dokładności lokalizacji danych w OSM potwierdzają także wartości współczynnika zmienności (v), wynoszące dla przesunięcia wzdłuż osi x ponad 125%, a osi y – prawie 170%. Aż 25% analizowanych danych charakteryzuje się większym od średniej (17,6 m) przesunięciem granicy w stosunku do ortofotomapy.

Wyniki wartości wariancji i odchylenia standardowego przesunięcia granicy wzdłuż osi x i y wykazują większe rozproszenie wyników wokół średniej wzdłuż osi x , na co może mieć wpływ cień drzew widoczny na ortofotomapie. Średni błąd lokalizacji lasów w OSM w stosunku do ortofotomapy wynosi 16,0 m, a błędy lokalizacji wzdłuż osi x i y odpowiednio 11,4 i 11,2 m (tab. 1).

Pochodzenie informuje potencjalnego użytkownika o materiałach źródłowych, na podstawie których pozyskano dane. Lokalizacja poszczególnych obiektów w OSM jest wprowadzana w wyniku pomiarów GPS lub wektoryzacji oficjalnie udostępnionych zasobów teledeteekcyjnych lub mapowych. Od września 2013 roku można za zgodą Głównego Geodety Kraju wykorzystywać dane udostępniane w portalu infrastruktury informacji przestrzennej w Polsce (geoportal.gov.pl). Wartości atrybutów określane są najczęściej na podstawie znajomości mapowanego terenu, ponieważ większość edytorów OSM kartuje obszary rodzinnych miast lub miejsca położone w ich sąsiedztwie [Marczak 2017]. Informacja o pochodzeniu nie jest zapisywana w żadnym z atrybutów danych OSM.

Kompletność danych OSM jest średnio o 14% mniejsza niż danych urzędowych, natomiast dokładność lokalizacji jest porównywalna z urzędowymi danymi topograficznymi, a w wielu przypadkach nawet lepsza (tab. 2). Dokładność danych BDOT10k i BDOO określona została jako przybliżona [Rozporządzenie... 2011]. Jeśli przyjąć, że wynosi ona 0,1-0,2 mm w skali opracowania, to wartości te wynoszą 10-20 m dla BDOT10k i 200-400 m dla BDOO. Otrzymana dokładność lokalizacji danych OSM wynosząca 16 m jest porównywalna z dokładnością lokalizacji lasów w Bazie Danych Obiektów Topograficznych (BDOT10k). Najmniej dokładnie wprowadzone kontury lasów (traktowane jako błędy grube) odpowiadają dokładności BDOO.

Dyskusja

Aktywne włączenie użytkowników internetu w pozyskiwanie danych przestrzennych sprawiło, że dane społecznościowe, szczególnie dane OSM, stanowią dynamiczne, stale aktualizujące się bazy danych, których ocena w pewnym sensie jest zawsze nieaktualna, bo dotyczy kopii pozyskanej w konkretnym czasie, a nie wszystkich danych dostępnych w danej chwili. Widać to

Tabela 1.

Średnia (M), mediana (Me), rozstęp (R), wariancja (S^2), odchylenie standardowe (SD), współczynnik zmienności (CV) i średni błąd kwadratowy (RMSE) wektorów przesunięcia granicy lasu

Mean (M), median (Me), range (R), variation (S^2), standard deviation (SD), coefficient of variability (CV) and root square mean error (RMSE) of forest border shift vectors

	Δx [m]	Δy [m]	Δd [m]
M	7,9	5,7	17,6
Me	6,8	4,8	11,4
R	76,6	77,0	342,0
S^2	98,45	94,08	825,89
SD	9,92	9,70	28,74
CV [%]	125,7	169,6	160,5
RMSE	11,4	11,2	16,0

Tabela 2.

Porównanie jakości danych OSM, BDOT10k i BDOO
Comparison of OSM, BDO10k and BDOO data quality

	OSM	BDOT10k	BDOO
Kompletność obiektów [%] Objects completeness	86,2	100 ±2	100 ±2
Kompletność atrybutów [%] Attributes completeness	94,2	100 ±2	100 ±2
Spójność pojęciowa [%] Logical consistency	100	100	100
Spójność formatu [%] Format consistency	100	100	100
Spójność topologiczna [%] Topological consistency	96,4	100	100
Dokładność lokalizacji Positional accuracy	16,0 m	przybliżona approximate	przybliżona approximate
Dostęp Availaility	ODbL	płatny z wyjątkiem administracji publicznej i uczelni; formaty gml i shp free only for public authorities and high schools; gml and shp formats	bezpłatny, format gml free, gml format
Pochodzenie Lineage	geoportal.gov.pl; obrazy satelitarne, ortofotomapy satellite imagery orthophotomaps	ortofotomapy, mapa zasadnicza, EGiB orthophotomaps basic map EGiB	BDOT10k

także podczas analizy danych dotyczących lasów. W pliku z obszaru województwa lubuskiego pozyskanym w styczniu 2018 roku powierzchnia lasów wynosi 631 714,23 ha, co powoduje, że kompletność zbioru wzrasta do 79,9% (o 3,5%). Proces ten jest niestety nieregularny i nie można przewidzieć, kiedy skartowane zostaną wszystkie obiekty terenowe [Neis, Zipf 2012; Marczak 2015]. Według danych OSM Polska z 2018 roku liczba edytorów OSM ciągle rośnie, a najbardziej aktywny twórca wprowadzał w 2017 roku dane niemal codziennie, głównie na terenie województwa zachodniopomorskiego (kompletność lasów wynosi tam prawie 90%). Do serwisu dodano także na obszarze kilku gmin dane pochodzące z rejestrów urzędowych, głównie adresy i budynki z EGiB, co znacząco podnosi użyteczność OSM. Analizując tempo, w jakim przyrastają dane OSM (10% w roku 2017), można zaryzykować stwierdzenie, że na pełną kompletność nie trzeba będzie długo czekać.

Kluczowe znaczenie dla określenia powierzchni lasów ma stosowana definicja. Różnice w definiowaniu lasu powodują, że powierzchnia obliczana na podstawie różnych źródeł nie jest wielkością stałą [Björk, Skånes 2015], a jej dokładne określenie nie jest możliwe [Hościło i in. 2016]. Definicja lasu stosowana w OSM zasadniczo nawiązuje do pokrycia terenu, determinując tym samym zakres zbieranych informacji.

Reprezentacją geometryczną lasów w bazie OSM jest wielobok określający ich zasięg, a podstawowym źródłem danych jest wektoryzacja ortoobrazów lub pomiar ręcznym odbiornikiem GPS. Na dokładność geometryczną danych wpływają nie tylko materiały źródłowe i techniki pozyskiwania, ale także umiejętność rozpoznania lasu i interpretacji przebiegu jego granicy. Goodchild [2007], promotor przestrzennych danych społecznościowych, stwierdził, że dobrze rozróżnialne obiekty geograficzne, m.in. lasy, są łatwo identyfikowane i dzięki temu częściej i dokładniej mapowane. Potwierdziły to także przedstawione badania kompletności i dokład-

ności lokalizacji lasów w OSM na terenie Polski. Kompletność lasów jest znacznie wyższa niż w roku 2014 – według Marczak [2015] wynosiła wtedy 56%, przy czym największą kompletnością wyróżniało się wówczas województwo lubuskie.

Średni błąd kwadratowy lokalizacji lasów w OSM względem ortofotomapy wynosi 16,0 m i jest porównywalny z dokładnością odczytania położenia lasu na mapie w skali 1:10 000. Nowak i in. [2005] zwracają także uwagę na aspekty kulturowe i prawne, które w dużym stopniu determinują sposób postrzegania rzeczywistości przez użytkowników danych.

Podsumowanie

Hipoteza badawcza mówiąca, że społecznościowe, ogólnodostępne dane o lasach dostarczane przez serwis OpenStreetMap są wiarygodnym źródłem danych, porównywalnym z danymi urzędowymi, została tylko częściowo potwierdzona. W zakresie kompletności obiektów dane OSM możemy uznać za w pełni użyteczne tylko dla trzech województw: opolskiego, zachodniopomorskiego i wielkopolskiego. W pozostałych niedomiar przekraczający 10% może znacznie wpłynąć na wyniki analiz. Nie możemy też liczyć na pozyskanie jakiegokolwiek charakterystyki lasów gromadzonych w OSM. Nazwy ma bardzo niewiele kompleksów leśnych, przy czym nie wszystkie są zgodne z nazwami oficjalnymi. Dokładność lokalizacji określona na 16 m odpowiada dokładności geometrycznej lasów zgromadzonych w BDOT10k i jest wielokrotnie wyższa od dokładności BDOO. Warto jednak pamiętać, że w zasobach OSM znajdują się także lasy, których zasięg został określony bardzo niedokładnie i może być porównywalny tylko do BDOO.

Mimo wspomnianych braków OSM stanowi cenne alternatywne źródło danych. Zbiory są łatwo dostępne, bezpłatne i bez ograniczenia możliwości ich wykorzystania, co niejednokrotnie czyni je konkurencyjnymi w stosunku do danych urzędowych. Wykorzystując je w różnego rodzaju analizach, należy jednak pamiętać o ciągle dużej ich heterogeniczności.

Literatura

- Arsanjani J. J., Bakillah M. 2015. Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap. *International Journal of Digital Earth* 8 (11): 861-876.
- Bielecka E. 2015. Geographical data sets fitness of use evaluation. *Geodetski vestnik* 59 (2): 335-348.
- Bielecka E., Medyńska-Gulij B. 2015. Zur Geodateninfrastruktur in Polen. *Kartographische Nachrichten* KN 4: 201-208.
- Björk A., Skånes H. 2015. The Need for Awareness of Semantic Plasticity in International Harmonization of Geographical Information: Seen from a Nordic Forest Classification Perspective. W: Ahlqvist O., Varanka D., Fritz, S., Janowicz K. [red.]. *Land Use and Land Cover Semantics*. Boca Raton, CRC Press.
- Budhathoki N. 2010. Participants' Motivations to Contribute to Geographic Information in an Online Community. University of Illinois, USA.
- Cichociński P. 2012. Ocena przydatności OpenStreetMap jako źródła danych dla analiz sieciowych. *Roczniki Geomatyki* 10 (7): 15-24.
- Girres J. F., Touya G. 2010. Quality assessment of the French OpenStreetMap dataset. *Transaction in GIS* 14 (4): 435-459.
- Goodchild M. F. 2007. Citizens as sensors: the Word of volunteered geography. *GeoJournal* 69 (4): 211-221.
- Haklay M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets, *Environment and Planning B. Planning and Design* 37: 682-703.
- Hościło A., Mironczuk A., Lewandowska A. 2016. Określenie rzeczywistej powierzchni lasów w Polsce na podstawie dostępnych danych przestrzennych. *Sylwan* 160 (8): 627-634.
- Izdebski W. 2017. Analysis of the cadastral data published in the Polish Spatial Data Infrastructure. *Geodesy and Cartography* 66 (2): 227-240. DOI: 10.1515/geocart-2017-0015.
- Jabłoński M. 2015a. Definicja lasu w ujęciu krajowym i międzynarodowym oraz jej znaczenie dla wielkości i zmian powierzchni lasów w Polsce. *Sylwan* 159 (6): 469-482.
- Jabłoński M. 2015b. Powierzchnia gruntów leśnych – przyczyny zmian i spójność źródeł danych. *Wiadomości Statystyczne* 60 (11): 54-68.

- Luc M., Bielecka E. 2015. Ontology for National Land Use / Land Cover Map: Poland Case Study. W: Ahlgvist O., Varanka D., Fritz S., Janowicz K. [red.]. Land Use and Land Cover Semantics: Principles, Best Practices, and Prospects. CRC Press. 21-40.
- Ludwig I., Voss A., Krause-Traudes M. 2011. A comparison of the street networks of Navteq and OSM in Germany. Adv. Geoinf. Sci. Chang. World 1: 65-84.
- Marczak S. 2015. Ocena zaangażowania społeczeństwa w tworzenie danych przestrzennych w Polsce na przykładzie projektu OpenStreetMap. Roczniki Geomatyki 13 (3): 239-253.
- Marczak S. 2017. Wyznaczenie cech społeczeństwa wpływających na zaangażowanie w tworzenie VGI w Polsce. Roczniki Geomatyki 15 (2): 233-248.
- Neis P., Zielstra D., Zipf A. 2013. Comparison of volunteered geographic information data contributions and community development for selected world regions. Future Internet 5 (2): 282-300.
- Neis P., Zipf A. 2012. Analyzing the contributor activity of a Volunteered Geographic Information project – The case of OpenStreetMap. ISPRS International Journal of Geo-Information 1 (2): 146-165.
- Noszczyk T., Hernik J. 2016. Kompleksowa modernizacja ewidencji gruntów i budynków. Acta Scientiarum Polonorum 15 (1): 3-17. DOI: 10.15576/ASP.FC/2016.15.1.
- Nowak Da Costa J. 2016. Novel tool for examination of data completeness based on a comparative study of VGI data and official building datasets. Geodetski vestnik 60 (3): 495-508.
- Nowak Da Costa J., Bielecka E., Całka B. 2015. Jakość danych OpenStreetMap – analiza informacji o budynkach na terenie Siedlecczyzny. Roczniki Geomatyki 14 (2): 201-211.
- Nowak J., Noureas-Iso J., Peedell S. 2005. Issues of multilinguality in creating a European SDI – the perspective for spatial data interoperability. W: Fullerton K. [red.]. Abstract book of 11th EC-GI&GIS Workshop – ESDI: Setting the Framework, European Commission. DG Joint Research Centre. 47-48.
- Rozporządzenie MSWiA z dnia 17 listopada 2011 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów ogólnogeograficznych, a także standardowych opracowań kartograficznych. 2011. Dz. U. Nr 279, poz. 1642.
- Schmidt M., Klettner S. 2013. Gender and experience-related motivators for contributing to openstreetmap. International workshop on action and interaction in volunteered geographic information (ACTIVITY), Leuven.
- Steinmann R., Häusler E., Klettner S., Schmidt M., Lin Y. 2013. Gender Dimensions in UGC and VGI: A Desk-Based Study. W: Jekel T., Car A., Griesebner G. [red.]. GI_Forum Creating the GISociety, Germany.
- Zielstra D., Zipf A. 2010. A comparative study of proprietary geodata and volunteered geographic information for Germany. 13th AGILE International Conference on Geographic Information Science 2010, Portugal.