

Ocena zaangażowania społeczeństwa w tworzenie danych przestrzennych w Polsce na przykładzie projektu OpenStreetMap

Assessment of society involvement in creation of spatial data
in Poland on the example of OpenStreetMap project

Sylwia Marczak

Wydział Geodezji i Kartografii, Politechnika Warszawska,
Zakład Fotogrametrii, Teledetekcji i Systemów Informacji Przestrzennej

Słowa kluczowe: społecznościowe dane przestrzenne, społeczeństwo informacyjne, system informacji geograficznej, dane topograficzne

Keywords: volunteered geographic information, information society, Geographic Information System, topographic data

Wstęp

Szybki rozwój technologii ma istotny wpływ na społeczeństwo, a można nawet stwierdzić, że bierze udział w jego kształtowaniu. Dotyczy to również technologii geoprzestrzennych, które niewątpliwie przyczyniają się do tworzenia społeczeństwa geoinformacyjnego. Jak podaje Gaździcki (2012) *publiczna świadomość w zakresie technologii geoprzestrzennych szybko się rozszerza*. Przy czym, ma to związek nie tylko z chęcią korzystania z danych przestrzennych, ale również aktywnego udziału w ich tworzeniu przez społeczeństwo. Należy podkreślić, że udział ten jest dobrowolny i działa na zasadzie wolontariatu. Dane w ten sposób tworzone określane są terminem *voluntereed geographic information (VGI)* sformułowanym przez Goodchild'a (2007) i do tej pory nie doczekały się tłumaczenia na język polski, dlatego w tym artykule będzie używany skrót *VGI*. Pozyskiwanie danych przestrzennych przez osoby niebędące profesjonalistami w tym zakresie stało się możliwe dzięki (Goodchild i in., 2012):

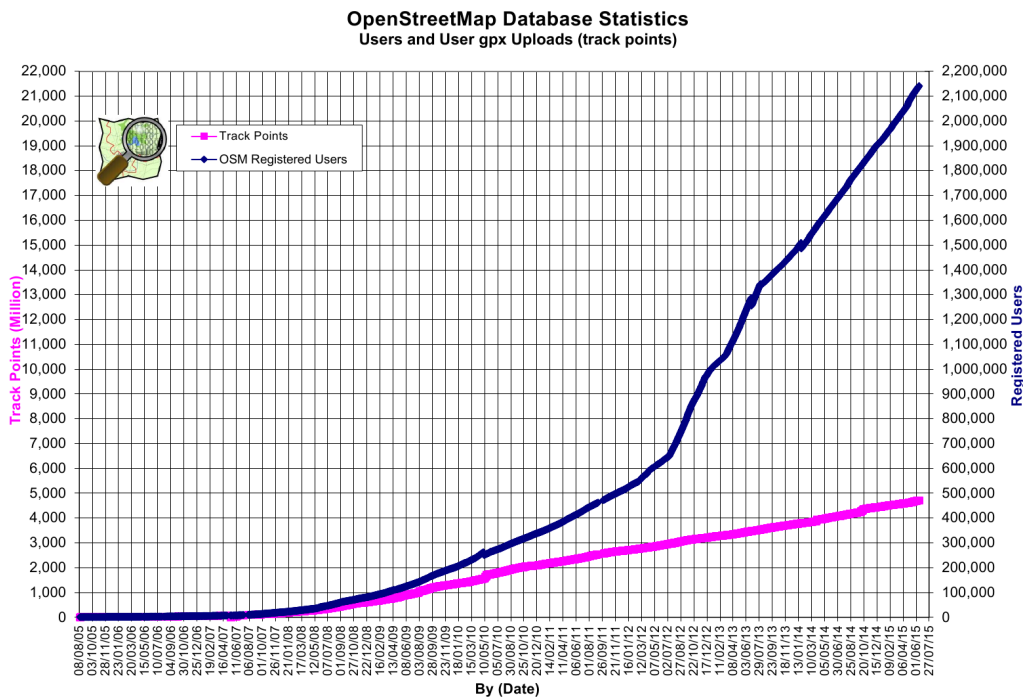
- technologii Web 2.0, umożliwiającej generowanie treści przez użytkowników serwisów internetowych,
- telefonom komórkowym z odbiornikami GPS (ang. *Global Positioning System*), zapewniającymi kilkunastometrową dokładność wyznaczania pozycji,
- systemom satelitarnym i naziemnym pozyskującym dane przestrzenne.

Termin *VGI* jest używany zamiennie z terminem *crowdsourcing* (ang. *crowd* – tłum, ang. *sourcing* – czerpanie), który został pierwszy raz użyty w 2006 roku przez dziennikarza J. Howe'a i oznacza *outsourcing* zadań wykonywanych w danej firmie lub instytucji do niezidentyfikowanej, zazwyczaj bardzo dużej grupy ludzi najczęściej za pośrednictwem Internetu. Przy czym należy tu podkreślić, że *crowdsourcing* w odniesieniu do przestrzeni geograficznej oznacza, że pozyskiwane dane mogą dotyczyć wszystkich zjawisk zachodzących na Ziemi, na przykład: pogody, katastrof, a nawet fenologii. Zjawisko to podnosi świadomość użytkowników dotyczącą wartości informacji przestrzennej, co ma bezpośredni wpływ na budowanie społeczeństwa geoinformacyjnego. Jest to szczególnie istotne w kontekście budowy krajowych infrastruktur informacji przestrzennych (*SDI* – ang. *Spatial Data Infrastructure*), w których niektórzy autorzy widzą również miejsce dla *VGI* (Goodchild, 2007; Craglia i in., 2008; Ho, Rajabifard, 2010; Castelein i in., 2010). Wśród zalet takiego podejścia wymieniane są: zwiększenie świadomości społecznej dotyczącej *SDI* dzięki popularności i modzie na *crowdsourcing*, pozyskiwanie dużej ilości bardzo aktualnych danych, a także włączenie społeczeństwa w proces podejmowania decyzji w oparciu o *SDI* (Craglia, 2007; Ho, Rajabifard, 2010). Literatura wskazuje *VGI* jako jeden z trendów technologii geoprzestrzennych (Goodchild i in., 2012; Gaździcki, 2012), co stanowiło motywację do powstania tego artykułu, którego celem jest zbadanie zaangażowania społeczeństwa w tworzenie danych przestrzennych w Polsce. Należy pamiętać, że w przypadku każdego rodzaju danych, a zwłaszcza w odniesieniu do *VGI* powstaje pytanie o ich wiarygodność, która do tej pory była badana w odniesieniu do danych przestrzennych wytworzonych przez profesjonalistów (m.in. Zielstra, Zipf, 2010; Hakley, 2010). Podejście to zostanie również zastosowane w artykule, którego drugim celem jest zbadanie w jakim stopniu dane *VGI* dla Polski podobne są do tych gromadzonych w ramach polskiej infrastruktury informacji przestrzennej.

OpenStreetMap jako przykład *VGI*

Rosnące zainteresowanie użytkowników Internetu nie tylko korzystaniem z informacji przestrzennej, ale również jej tworzeniem sprawiło, że liczba serwisów oferujących taką funkcjonalność ciągle rośnie. Jednym z najbardziej popularnych przykładów *crowdsourcingu* jest projekt OpenStreetMap (OSM), który został rozpoczęty w 2004 roku przez Steve'a Coasts'a (Zielstra, Zipf, 2010). W swojej początkowej fazie miał dotyczyć stworzenia możliwie najdokładniejszej mapy Wielkiej Brytanii przez użytkowników Internetu. Główną motywacją powstania projektu była chęć darmowego udostępnienia zebranych informacji i możliwość ich wykorzystania do dowolnego celu (Haklay, 2008). Wynikła ona z wysokich cen danych przestrzennych oferowanych przez firmy komercyjne i agencje rządowe zajmujące się ich pozyskiwaniem, co było i jest nadal problemem znanym również w Polsce. Szybki wzrost popularności serwisu wyrażony liczbą zarejestrowanych użytkowników (rys. 1) sprawił, że projekt został rozszerzony na cały świat, umożliwiając każdemu korzystającemu z Internetu tworzenie *VGI*. Dane można edytować bezpośrednio na stronie WWW projektu lub wykorzystując zewnętrzne narzędzia *of-line*. Można się przy tym wspomóc zebranymi przez siebie pomiarami GPS lub wykonać wektoryzację z wykorzystaniem ortofotomapy udostępnione przez firmy Yahoo i Microsoft. Wytworzone przez wolontariuszy dane przestrzenne są rozpowszechniane na licencji Open Database Licence (ODbL), a to oznacza, że mogą być pobierane, edytowane i ponownie wykorzystywane pod warunkiem podania OpenStreetMap

jako źródła. Ciągłe rosnąca popularność projektu, przekładająca się na liczbę pozyskiwanych danych przestrzennych, skłania do zastanowienia się nad przyczynami zaangażowania użytkowników Internetu w tworzenie *VGI*. Jak podają Kukułka i Gotlib (2014) mogą to być: chęć promowania samego siebie, budowania relacji ze społecznością oraz potrzeba udoskonalenia istniejących danych według własnych potrzeb. Według Coleman'a i in. (2009) motywacją do tworzenia *VGI* może być również altruizm, zainteresowania danej osoby, chęć rozwoju intelektualnego lub nawet lokalny patriotyzm. Współczesne możliwości technologii mobilnych, dostęp do Internetu i odbiorników GPS sprawiają, że pozyskiwanie *VGI* jest łatwe, a poczucie tworzenia czegoś dla społeczeństwa może przyczyniać się do odczuwania przyjemności i satysfakcji. Powody te wpłynęły i nadal wpływają na rosnącą liczbę użytkowników OSM na całym świecie, która obecnie wynosi ponad 2 miliony (rys. 1). Powstaje jednak pytanie, jak dużym potencjałem dysponują w kontekście tworzenia *VGI* i jak go mierzyć. Próba odpowiedzi na nie jest zaprezentowana poniżej metodyka i otrzymane wyniki.



Rysunek 1. Liczba użytkowników OpenStreetMap i tworzonych przez nich danych przestrzennych (źródło: <http://wiki.openstreetmap.org/wiki/Statistics>)

Wykorzystane dane przestrzenne – amatorskie i profesjonalne

Dane z projektu OpenStreetMap można pozyskać na kilka sposobów. Można to zrobić bezpośrednio na stronie WWW przez zaznaczenie obszaru i zapisanie treści topograficznej w postaci plików .xml. Można również skorzystać z narzędzi przeznaczonych do pobierania danych z OSM dostępnych w różnego rodzaju programach typu GIS. Natomiast sposobem,

z którego skorzystano w badaniach było pobranie danych ze strony WWW firmy Geofabrik w plikach .shp, które są udostępniane w podziale na państwa i codziennie aktualizowane. W swojej ofercie firma ma możliwość pobrania darmowych plików o ograniczonym zakresie tematycznym w plikach .shp i danych „surowych” w formacie .osm xml i .pbx oraz danych płatnych przetworzonych do struktury ułatwiającej korzystanie z nich. W strukturze bazy danych OSM wyróżniono trzy główne elementy: węzeł (*node*) – punkt o przypisanych współrzędnych geograficznych, linia (*way*) – zbiór węzłów służących do reprezentacji obiektów liniowych i poligonowych oraz relacja (*relation*), która określa zależności między elementami. Ta specyficzna budowa sprawia, że eksport do plików .shp może generować błędy w danych, które można jednak wyeliminować w programach typu GIS, a co za tym idzie umożliwić ich wykorzystanie w tym formacie. Należy jednak zwrócić uwagę, że w artykule analizowano w pewien sposób przetworzone dane OSM i wyciągnięte na ich podstawie wnioski mogą być podstawą jedynie do wykrycia ogólnych tendencji dotyczących zaangażowania społeczeństwa w ich tworzenie.

W celu zbadania aktywności użytkowników i jakości tworzonych przez nich danych skorzystano z danych darmowych pobranych w trzech punktach czasu – 01.2014 r., 05.2014 i 05.2015 r., co pozwoliło również na wyznaczenie dynamiki pozyskiwania treści przez użytkowników OSM. W ramach darmowych danych dostępnych jest osiem warstw:

- punktowe – *Points* (punkty zainteresowania) i *Places* (miejsowości);
- liniowe – *Railways* (linie kolejowe), *Waterways* (osie liniowych wód powierzchniowych) i *Roads* (przebiegi dróg);
- powierzchniowe – *Landuse* (pokrycie terenu), *Natural* (obiekty naturalne) i *Buildings* (budynki).

W dalszej części artykułu posłużono się anglojęzycznymi nazwami warstw OSM.

Ze względu na fakt, że OpenStreetMap nie ma przypisanej szczegółowości treści i dokładności geometrycznej, tak jak ma to miejsce w przypadku krajowych bazodanowych zbiorów danych przestrzennych, wchodzących w skład Krajowej Infrastruktury Informacji Przestrzennej (KIIP), podjęto próbę porównania ich (tab. 1) z danymi małoskalowymi i

Tabela 1. Klasy obiektów ze zbiorów wchodzących w skład KIIP i warstwy z OSM wybrane do porównania *VGI* i profesjonalnych danych przestrzennych

Nazwa obiektu w danych KIIP	Nazwa klasy obiektów	Nazwa warstwy OSM	Sposób wyboru obiektów z warstwy OSM	Geometria danych
Porównanie OSM z BDOO				
Las, zagajnik i zadrzewienie	PTLZ_A	<i>Natural</i>	TYPE = forest	poligonowe
Tor kolejowy	SKTR_L	<i>Railways</i>	TYPE = rail	liniowe
Miejscowość	ADMS_P	<i>Places</i>	TYPE = town, city, village	punktowe
Porównanie OSM z BDOT10k				
Budynek Budowla sportowa	BUBD_A BUSP_A	<i>Buildings</i>	–	poligonowe
Jezdnia	SKJZ_L	<i>Roads</i>	TYPE = living_street, motorway, primary, residential, secondary, tertiary	liniowe
Przystanek autobusowy lub tramwajowy	OIKM_P (kod obiektu OIKM04)	<i>Points</i>	TYPE = bus_stop, tram_stop	punktowe

wielkoskalowymi w celu kompleksowego przeanalizowania podobieństwa *VGI* do danych profesjonalnych. Skorzystano z Bazy Danych Obiektów Ogólnogeograficznych (BDOO), która odpowiada opracowaniom analogowym w skali 1:250 000 – posłużyła do porównań na poziomie regionalnym, a także z Bazy Danych Obiektów Topograficznych (BDOT10k) odpowiadającej skali 1:10 000 – do porównań na poziomie lokalnym. Aktualność geometryczna i tematyczna danych z BDOO to 31.12.2013 roku, natomiast dla BDOT10k są one zróżnicowane i wahają się w zależności od klasy obiektów w przedziale od 10.03.2009 do 31.12.2013 roku.

Wykorzystano po trzy klasy obiektów, o każdym typie geometrycznym z obu baz danych, w celu porównania z odpowiadającymi obiektami z OSM, co najczęściej wiązało się z ich wcześniejszą selekcją za pomocą atrybutu Type (typ obiektu). Warstwy wybrano w taki sposób aby miały swoje odpowiedniki w OSM, przy czym do porównania z BDOO wybrano te, na których znajdują się obiekty o zasięgu krajowym (las, linie kolejowe, miejscowości), natomiast do porównania z BDOT te o zasięgu lokalnym (budynki, jezdnie, przystanki) (tab. 1).

Metody analizy i porównania *VGI* i profesjonalnych danych przestrzennych

Celem badań było zbadanie zaangażowania użytkowników OSM w tworzenie danych przestrzennych, a także porównanie ich z profesjonalnie pozyskanymi danymi przestrzennymi, w kontekście oceny podobieństwa *VGI* do danych wchodzących w skład KIIP. Należy przy tym podkreślić, że wszystkie niżej opisane i wykonane analizy przedstawiają jedynie porównanie między zbiorami OSM w czasie, bądź danymi z BDOO i BDOT10k i nie mają odniesienia do świata rzeczywistego, ani do zmian jakie w nim zachodzą. Porównania OSM dokonano względem innych danych przestrzennych, a co za tym idzie otrzymane wyniki nie świadczą o bezwzględnej jakości tego zbioru.

Przed przystąpieniem do analiz sformułowano następujące założenie – o zaangażowaniu użytkowników OSM w tworzenie danych przestrzennych świadczy wzrost liczby danych w tym zbiorze w czasie, który obarczony jest zaniedbywalnym wpływem zmian, które zaszły w świecie rzeczywistym dzięki krótkim przedziałom czasu pozyskania kolejnych wersji OSM. Ponadto, dzięki przeprowadzeniu analiz w odniesieniu do jednostek podziału terytorialnego kraju możliwe będzie wykrycie regionalnych i lokalnych zależności dotyczących zaangażowania użytkowników w tworzenie *VGI* w Polsce. Porównanie OSM z krajowymi zbiorami danych przestrzennych mimo tego, że charakteryzują się różną aktualnością, pozwoli na zbadanie potencjału *VGI* do aktualizacji tych zbiorów wraz ze wskazaniem klas obiektów, dla których ma to największe uzasadnienie.

Zaangażowanie użytkowników w tworzenie danych OSM badano za pomocą wzrostu jakości części opisowej danych oraz wzrostu liczby danych, wyrażonego przyrostem rozmiarów części geometrycznej poszczególnych warstw – dla warstw poligonowych była to powierzchnia wyszczególnionych obiektów, dla liniowych – długość, a dla punktowych – liczba obiektów. Porównano liczbę atrybutów poszczególnych warstw dla trzech okresów, w których pozyskano dane, a także stopień ich wypełnienia. W przypadku porównania części geometrycznej odczytano sumy rozmiarów obiektów na poszczególnych warstwach i obliczono dynamikę zmian dla każdej z nich dla całej Polski, a także w podziale na powiaty.

Porównania danych OSM z bazami danych z KIIP dokonano w zależności od typu geometrycznego danych w oprogramowaniu ArcGIS, z wykorzystaniem wbudowanych i własnych narzędzi. Do porównania danych powierzchniowych wykorzystano indeks Jaccard'a, który jest miarą podobieństwa pomiędzy dwoma zbiorami i wyraża się następującym wzorem (Bakuła, 2014):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad \text{gdzie } A, B - \text{zbiory między którymi badane jest podobieństwo}$$

Wartości indeksu mieszczą się w przedziale [0;1], w którym wartość 0 oznacza całkowitą rozłączność zbiorów, a co za tym idzie brak podobieństwa, natomiast wartość 1 – tożsamość zbiorów. Za moce części wspólnej i sumy zbiorów przyjęto powierzchnie geometrycznej sumy i części wspólnej porównywanych klas obiektów z BDOO i BDOT10k oraz odpowiadających im danych z OSM.

Porównania danych liniowych dokonano wyznaczając części wspólne warstw liniowych, a także części uwzględnione tylko w jednym z porównywanych zbiorów. Za tożsame uznawano te obiekty liniowe z porównywanych zbiorów, których przebiegi nie różniły się między sobą o więcej niż 10 metrów. W wyniku otrzymano procent danych OSM pokrywających się z bazami BDOO i BDOT10k, a także części zbiorów, które zostały wyszczególnione tylko w OSM oraz tylko w danych KIIP.

W celu porównania warstw punktowych stworzono narzędzie, które wyszukiwało najbliższe położone względem siebie obiekty na porównywanych warstwach – jednej z OSM i jednej z BDOO lub BDOT10k. Następnie porównywało nazwy obiektów i zliczało liczbę tych, które sobie odpowiadały. Pozwoliło to na obliczenie procentu odpowiadających sobie nazw obiektów, wskazującego na podobieństwo danych opisowych porównywanych zbiorów. Następnie obliczono średnią odległość najbliższych położonych obiektów o odpowiadających nazwach, jako wskaźnika podobieństwa położenia danych z OSM względem BDOO lub BDOT10k.

Otrzymane wyniki

Wyniki otrzymane z porównania pozyskanych w trzech okresach zbiorów OSM świadczą o zdecydowanym wzroście treści tego produktu, a co za tym idzie wzroście zaangażowania społeczeństwa w tworzenie *VGI*. Ze względu na fakt, że analizowane przedziały czasowe były krótkie (5 miesięcy i rok) wpływ zmian topograficznych, które zaszły w rzeczywistości, uznano za niewielki i nie był brany pod uwagę w analizach. W największym stopniu wzrost liczby danych w czasie i ich szczegółowość związane są z aktywnością jego użytkowników. O szczegółowości danych gromadzonych w OSM stanowią: liczba warstw i atrybutów je opisujących oraz jakość ich wypełnienia. Wszystkie warstwy opisane są tym samym zestawem 3 atrybutów – identyfikator (*OSM_ID*), nazwa (*NAME*) i typ obiektu (*TYPE*). Warstwy *Places*, *Waterways*, *Roads*, *Points* mają dodatkowe atrybuty związane z charakterem gromadzonych danych (tab. 2). Jednak dla zdecydowanej większości obiektów atrybuty te nie są wypełnione, a co za tym idzie nieużyteczne (tab. 2). Liczba atrybutów w badanym okresie nie uległa zmianie, gdyż wiązałoby się to ze zmianą modelu danych. Dla większości warstw wzrosła liczba unikalnych wartości dla atrybutu *TYPE*, przy czym największy wzrost nastąpił dla warstwy *Points* (tab. 2). Niestety, wzrost ten nie wynika z uszczegóło-

Tabela 2. Liczba wartości dla atrybutu TYPE oraz stopień wypełnienia dodatkowych atrybutów poszczególnych warstw OSM

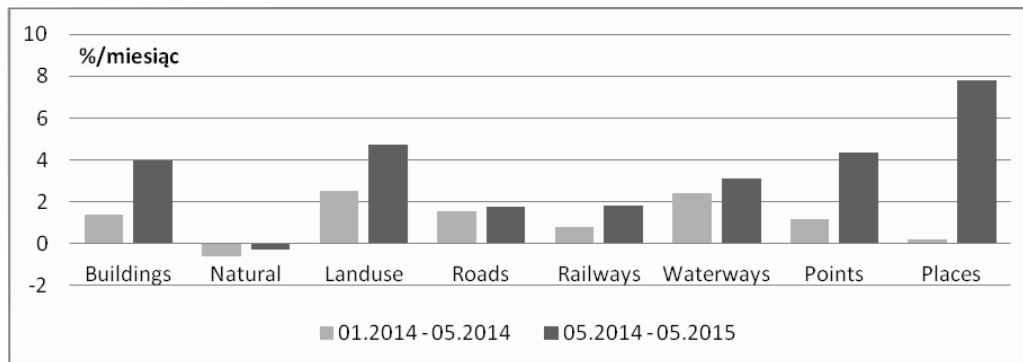
Nazwa warstwy	Liczba wartości dla atrybutu TYPE			Dodatkowe atrybuty nazwa	Stopień wypełnienia [%]		
	OSM 01.2014 r.	OSM 05.2014 r.	OSM 05.2015 r.		OSM 01.2014 r.	OSM 05.2014 r.	OSM 05.2015 r.
<i>Waterways</i>	30	36	33	WIDTH	1,03	0,87	0,75
<i>Places</i>	19	19	23	POPULATION	18,18	18,04	9,17
<i>Landuse</i>	117	124	141	-			
<i>Buildings</i>	434	449	551	TYPE*	21,86	21,94	19,85
<i>Roads</i>	45	58	64	ONEWAY BRIDGE TUNNEL MAXSPEED	5,16**	5,38**	6,24**
<i>Points</i>	824	838	1010	-			

*TYPE jest atrybutem podstawowym, przedstawiono go w celach porównawczych.

** Stopień wypełnienia atrybutu MAXSPEED.

wienia danych, a z błędów popełnianych przez użytkowników w momencie ich wprowadzania. Z ogólnej analizy wartości atrybutu TYPE dla wszystkich warstw wynika, że błędy wynikają w głównej mierze z: błędnej pisowni, braku wypełnienia oraz wpisywania nazwy obiektu, zamiast jego typu. Błędy te w dużym stopniu mogą być spowodowane niezbyt udaną próbą konwersji danych OSM zapisanych w plikach .osm xml charakteryzujących się specyficzną budową bazy danych na pliki .shp.

Analizując rozmiary części geometrycznej poszczególnych warstw OSM można stwierdzić znaczny miesięczny wzrost liczby danych w tym zbiorze w badanych przedziałach czasu (rys. 2). Tylko dla danych warstwy *Natural* zaobserwowano spadek powierzchni wyszczególnionych obiektów. Zarówno dla krótszego jak i dłuższego przedziału czasu można zauważyć wzrost prędkości przyrostu danych OSM. Analiza tylko dwóch okresów nie



Rysunek 2. Miesięczny przyrost danych OSM w procentach w Polsce w okresie 01.2014 do 05.2015 roku

pozwała stwierdzić istnienia trendu zmian, ale jest dowodem na wzrost zainteresowania społeczeństwa tworzeniem danych przestrzennych w ostatnim czasie.

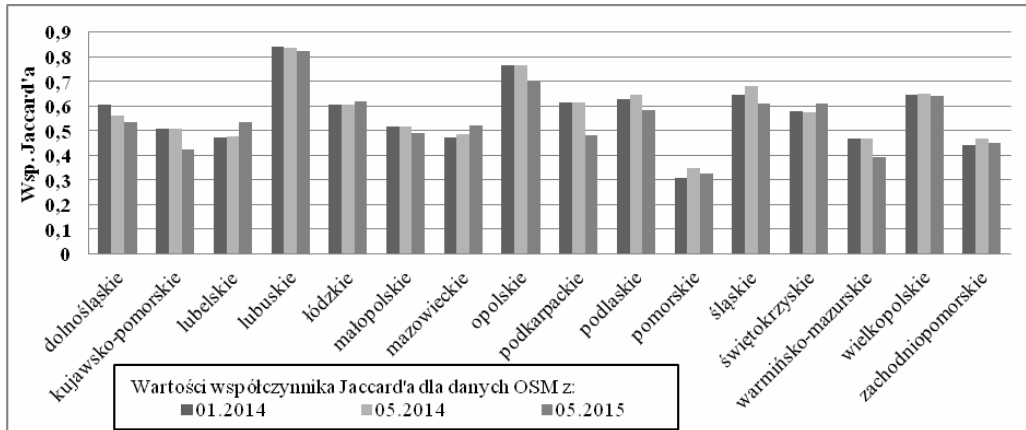
Wyniki otrzymane w podziale na powiaty, w celu wykrycia regionalnych zależności przestrzennych, najlepiej przedstawić w formie kartogramu. Jednak ze względu na ograniczenia wydawnicze sporządzono je tylko dla warstw *Places* i *Landuse*, gdyż dla nich nastąpił największy wzrost liczby danych. Wyniki otrzymane dla warstwy *Landuse* porównano z różnicą geometryczną warstw pozyskanych z danych z maja 2014 roku i maja 2015 roku (rys. 3). Można zauważyć że największy wzrost ilości danych nastąpił w powiatach północnej i północno-wschodniej Polski, natomiast najmniejszy w Polsce środkowej i południowo-zachodniej.

Wyniki otrzymane dla warstwy *Places* pozwalają na jednoznaczne zidentyfikowanie powiatów, w których nastąpił największy wzrost liczby miejscowości gromadzonych w OSM (rys. 4b). W danych z maja 2014 roku na warstwie wyróżniono 61 202 obiekty, natomiast rok później aż 123 357. Jednak trudno stwierdzić czy ma to związek z rzeczywistą liczbą miejscowości w południowej Polsce, czy raczej z nadaktywnością użytkowników z tego regionu. W związku z powyższym porównano otrzymane wyniki z liczbą miejscowości w powiatach zawartą w Systemie Identyfikatorów i Nazw Miejscowości (SIMC), będącym składnikiem Krajowego Rejestru Urzędowego Podziału Terytorialnego Kraju (TERYT), w którym zgodnie ze stanem na dzień 01.01.2015 roku wyszczególniono 103 124 miejscowości. Dla większości powiatów liczba miejscowości w OSM jest większa niż w SIMC (rys. 4a). W powiatach w których nastąpił znaczny wzrost ilości danych w okresie 05.2014 r.-05.2015 r. występują również duże rozbieżności względem rejestru SIMC (rys. 4). Zależność ta świadczy z jednej strony o dużej aktywności użytkowników, jednak z drugiej o niewystarczającym zwracaniu uwagi na jakość i poprawność wprowadzanych danych, a także błędach interpretacyjnych. Wykonana analiza ilościowa, może być wykorzystana do wskazania tych miejsc, gdzie dane wprowadzone do OSM należałoby poddać dokładniejszej analizie jakościowej uwzględniającej między innymi wypełnienie atrybutu *NAME*.

Wyniki otrzymane z porównania OSM do zbiorów wchodzących w skład KIIP wskazują na stosunkowo wysokie podobieństwo części geometrycznej. Ze względu na ograniczoną szczegółowość danych OSM, część opisowa była porównywana w ograniczonym zakresie. Analiz dokonano na dwóch poziomach – małoskalowym z wykorzystaniem BDOO dla całej Polski i w podziale na województwa oraz wielkoskalowym z wykorzystaniem BDOT10k dla dwóch gmin. Porównania dokonano dla sześciu warstw przedstawionych w tabeli 1.

Porównanie warstwy *PTLZ_A* i lasów pozyskanych z warstwy *Natural* z OSM wykonano w oparciu o współczynnik Jaccarda, którego wartość dla całej Polski wyniosła 0,56 dla danych OSM z 01.2014 roku, 0,56 – z 05.2014 roku i 0,53 dla danych z 05.2015 roku. Można zatem stwierdzić, że w badanym okresie podobieństwo tych dwóch zbiorów jest na porównywalnym poziomie i średnio wynosi 55%. Chcąc stwierdzić czy istnieją tendencje regionalne w zakresie podobieństwa badanych zbiorów obliczono współczynniki Jaccarda również dla województw (rys. 6). Największe podobieństwo występuje dla województwa lubuskiego, natomiast najmniejsze dla pomorskiego. Ponadto dla 4 województw: mazowieckiego, łódzkiego, świętokrzyskiego i lubelskiego – zaobserwowano wzrost podobieństwa badanych zbiorów.

Kolejną porównywaną warstwą były linie kolejowe – warstwa *Railways* z OSM i klasa obiektów *SKTR_L* z BDOO. Podobieństwo warstw wyrażone częścią wspólną linii kolejowych rośnie w czasie, jest wyższe niż w przypadku lasów i wynosi średnio dla całej Polski



Rysunek 6. Wartości współczynnika Jaccard'a dla klasy obiektów PTLZ_A i odpowiadających jej obiektów z warstwy *Natural* dla trzech okresów pozyskania danych OSM w podziale na województwa

61,6% (60,7% dla danych z 01.2014 roku, 61,8% – 05.2014 roku i 62,7% – 05.2015 roku). Ponadto rośnie długość linii kolejowych wyszczególnionych w OSM, ale nieposiadających swoich odpowiedników w BDOO z 25,7% na początku 2014 roku do 28,5% w połowie roku 2015. W ujęciu regionalnym można zauważyć, że na południu i południowym wschodzie Polski więcej jest linii kolejowych wyszczególnionych tylko w OSM niż tych wskazanych tylko w BDOO, co może być związane z większą aktywnością użytkowników tych regionów (rys. 5). Odwrotna sytuacja zachodzi na północy i północnym zachodzie kraju.

Ostatnie porównanie BDOO i OSM dotyczyło warstw punktowych reprezentujących miejscowości. Analizy dokonano w podziale na miasta i wsie, a otrzymane wyniki wskazują na bardzo wysoki stopień podobieństwa tych zbiorów (tab. 3). Obliczono w jakim stopniu nazwy miejscowości w obu zbiorach sobie odpowiadają oraz jaka jest średnia odległość odpowiadających sobie obiektów, jako wskaźnik podobieństwa ich położenia. Dla miast podobieństwo nazw wynosi średnio 95% natomiast dla wsi jedynie 78%. Miasta, reprezentowane przez punkty, charakteryzują się niższym stopniem podobieństwa położenia niż wsie

Tabela 3. Porównanie klasy obiektów ADMS_P z BDOO i warstwy Places z OSM

	Liczba obiektów			
	BDOO	OSM 01.2014 r.	OSM 05.2014 r.	OSM 05.2015 r.
Miasto	964	959	956	962
Wieś	43 349	48 031	48 306	50 036
Odpowiadające sobie nazwy miast w zbiorach BDOO i OSM [%]		94,8	94,9	94,3
Średnia odległość między odpowiadającymi sobie miastami [m]		406,9	406,0	402,8
Odpowiadające sobie nazwy wsi w zbiorach BDOO i OSM [%]		76,4	76,6	80,4
Średnia odległość między odpowiadającymi sobie wsiami [m]		300,0	298,6	287,3

(tab. 3). Jest to związane z faktem, że zajmują one duże obszary, co w połączeniu z dowolnością wskazania punktu je reprezentującego generuje większe rozbieżności. Zarówno w danych BDOO jak i OSM liczba miast nie jest zgodna z podawaną przez GUS (Główny Urząd Statystyczny) – według danych na dzień 01.01.2014 roku wynosiła ona 913, a rok później 915.

Z analizy zależności regionalnych wynika, że największe podobieństwo zbiorów w odniesieniu do miast występuje dla województwa świętokrzyskiego, natomiast dla wsi – opolskiego. Należy również podkreślić fakt, że w województwach, w których nastąpił wyjątkowo znaczny wzrost liczby miejscowości w OSM, czyli podkarpackiego i małopolskiego, średnia odległość odpowiadających sobie obiektów jest wysoka i wynosi 501,6 i 330,7 m. Dla województw pomorskiego i zachodniopomorskiego zaobserwowano najniższe podobieństwo nazw wsi z OSM i BDOO wynoszące zaledwie 60% i 64 %.

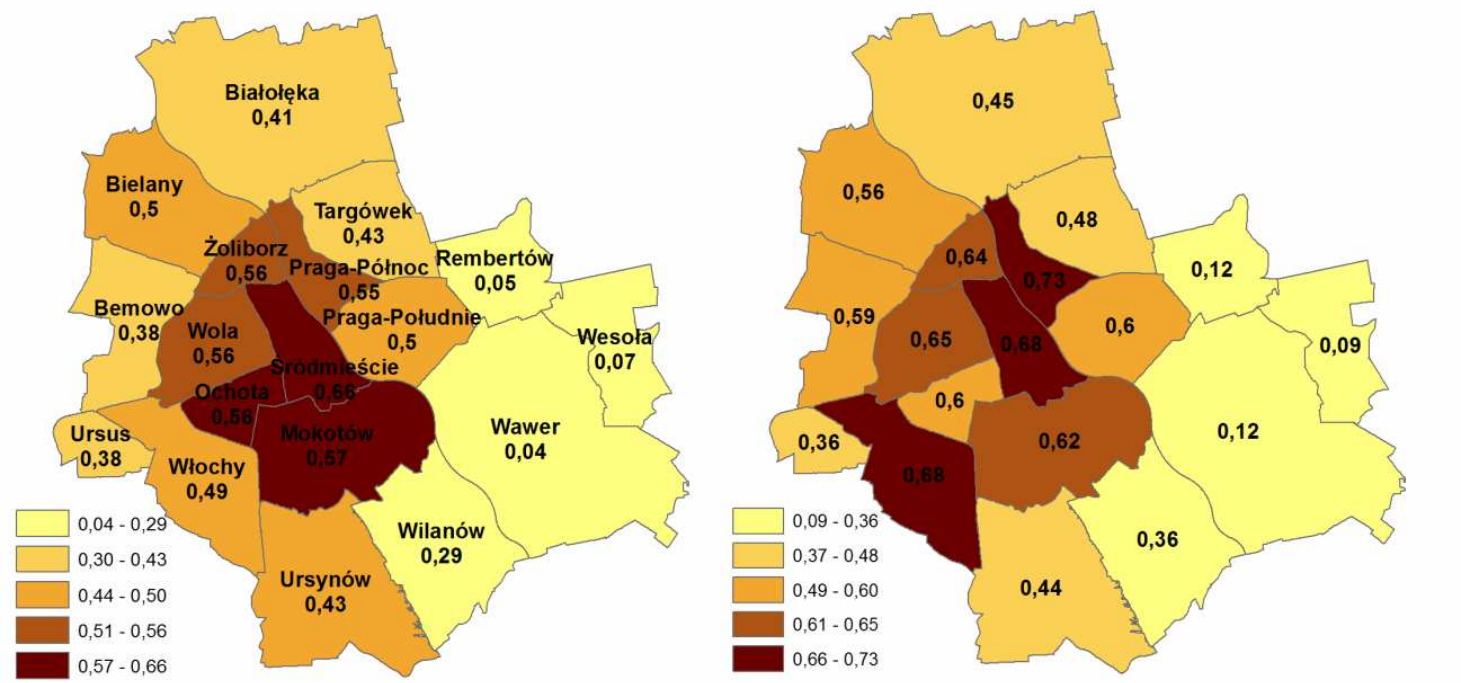
W przypadku danych z BDOT10k dokonano porównania dla dwóch gmin – Świeradów-Zdrój (gmina miejska w województwie dolnośląskim) i m.st. Warszawa. W przypadku stolicy Polski zastosowano podział na dzielnice w celu uszczegółowienia analiz. Aby określić podobieństwo warstwy budynków z BDOT10k i OSM po raz kolejny posłużono się indeksem Jaccard'a, którego wartość wyniosła średnio 46% dla Warszawy i zaledwie 7% dla Świeradowa-Zdroju. Można zatem stwierdzić, że dla dużych miast dane OSM są pozyskiwane przez użytkowników chętniej, częściej i z większą starannością.

Ponadto z analizy współczynnika Jaccard'a dla dzielnic Warszawy wynika, że najwyższe wartości są osiągane w centrum miasta i maleją wraz z oddalaniem się od niego, przy czym kierunek oddalania ma wpływ na wartość współczynnika (rys. 7). Jest on wyjątkowo niski dla dzielnic wschodnich, nieco wyższy na północy i południu miasta, a najwyższy na zachodzie (rys. 7). Jednak dla wszystkich dzielnic można zaobserwować jego wzrost, co świadczy o rosnącym w czasie podobieństwie do danych z BDOT10k.

Kolejną analizę wykonano dla przebiegów dróg. Podobieństwo wyrażone tak jak w przypadku kolei, częścią wspólną dla danych z BDOT10k i kolejnych wersji OSM wynosi średnio 63,9% dla Warszawy i zaledwie 30,9% dla gminy Świeradów-Zdrój. Należy jednak zauważyć, że w tej ostatniej nastąpił gwałtowny wzrost części wspólnej przebiegu dróg w okresie 05.2014–05.2015 roku, co nie miało miejsca w przypadku Warszawy i może mieć związek ze wzrostem aktywności użytkowników tamtego regionu (rys. 8). Ponadto, w przypadku obu gmin wzrasta liczba obiektów wyszczególnionych w OSM, a niemających swoich odpowiedników w BDOT10k, przy jednoczesnym wzroście części wspólnej obu zbiorów.

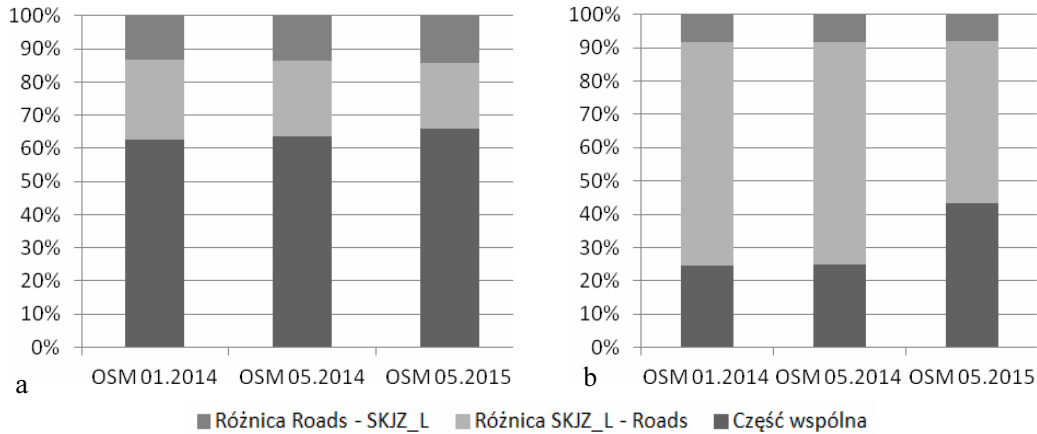
Analizując dzielnice Warszawy stwierdzono, że najmniejsze podobieństwo do danych BDOT10k występuje dla Wesołej, Rembertowa i Wawra – 55,1, 55,4 i 60,7%, natomiast największe dla Ursusa i Śródmieścia – 78,1 i 72,7%. Rozkład przestrzenny dzielnic o największym i najmniejszym podobieństwie jest bardzo zbliżony dla tego otrzymanego dla obrysów budynków i wskazuje te jednostki, jako te o najwyższej aktywności użytkowników OSM.

Ostatnim porównaniem wykonanym w odniesieniu do danych z BDOT10k była analiza podobieństwa położenia oraz nazw przystanków tramwajowych i autobusowych. W danych OSM informacje o tych obiektach można pozyskać z warstwy *Points*. Warstwa zawiera również dane, które w większości nie są gromadzone w zbiorach wchodzących w skład krajowej infrastruktury informacji przestrzennej (bankomaty, restauracje, sklepy wraz z nazwami, nazwy punktów usługowych), które jednocześnie są niezwykle atrakcyjne dla użytkowników OSM. Ze względu na fakt, że w gminie Świeradów Zdrój w danych OSM nie wyróżniono przystanków autobusowych, mimo ich istnienia według danych z BDOT10k, analizę przeprowadzono jedynie dla Warszawy.



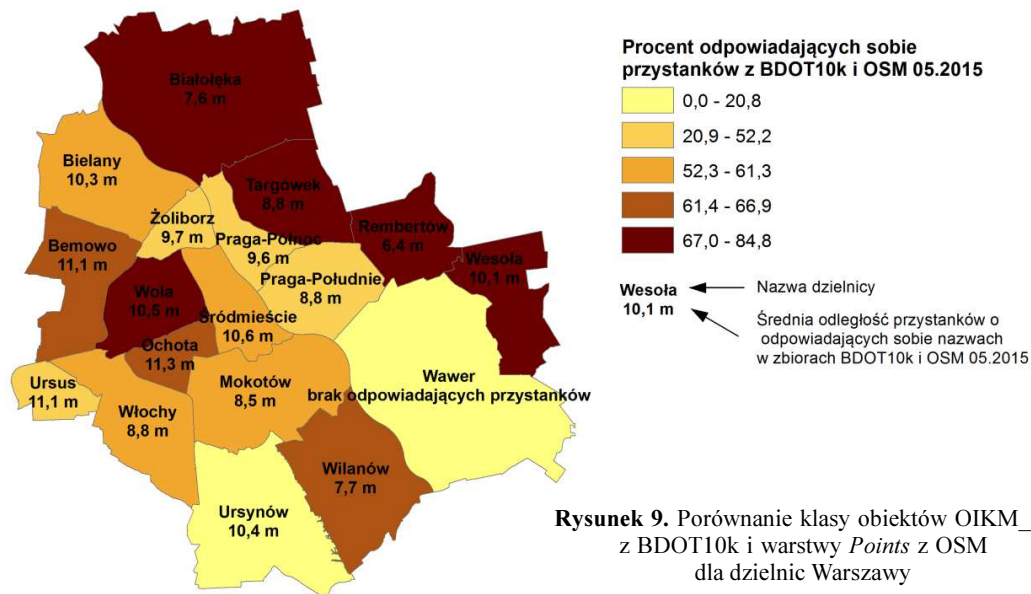
Współczynnik Jaccarda dla obrysów budynków z BDOT10k i OSM 01.2014 Współczynnik Jaccarda dla obrysów budynków z BDOT10k i OSM 05.2015

Rysunek. 7. Porównanie wartości współczynnika Jaccard'a dla obrysów budynków z BDOT10k i OSM dla dzielnic Warszawy



Rysunek 8. Porównanie klasy obiektów SKJZ_L i warstwy Roads dla gminy: a – m. st. Warszawa, b – Świeradów-Zdrój

W BDOT10k wyróżniono 3855 obiektów o kodzie OIKM04 – przystanek autobusowy lub tramwajowy. W OSM są one rozdzielone na przystanki autobusowe, których w 01.2014 roku było 3209 i tramwajowe – 552. Ich liczby w 05.2015 roku wynoszą odpowiednio – 3528 i 574, a co za tym idzie sumarycznie przewyższają liczbę tych wyszczególnionych w BDOT10k. Podobieństwo danych zostało zbadane w sposób analogiczny do zastosowanego w przypadku miejscowości, a więc polegało na sprawdzeniu w jakim procencie nazwy przystanków z obu zbiorów sobie odpowiadają i jaka jest średnia odległość przystanków o odpowiadających sobie nazwach. Otrzymane wyniki wskazują, iż zarówno podobieństwo nazw jak i położenia rosną nieznacznie w badanym okresie. Dla całego miasta nazwy przystanków odpowiadały sobie w 55,9% w 01.2014 i 05.2014 roku oraz 57% w 05.2015 roku, natomiast



Rysunek 9. Porównanie klasy obiektów OIKM_P z BDOT10k i warstwy Points z OSM dla dzielnic Warszawy

średnie odległości wynosiły ok. 12 metrów dla danych z 01.2014 i 05.2015 roku oraz 9 metrów w 05.2015 roku.

Z analizy zależności lokalnych zachodzących w Warszawie wynika, że największe podobieństwo między BDOT10k i OSM w zakresie przystanków autobusowych i tramwajowych zachodzi dla dzielnic Rembertów i Białołęka, natomiast najmniejsze dla Wawra i Ursynowa (rys. 9). Obserwacja ta nie jest zgodna z poprzednimi analizami, z których wynikało że największe podobieństwo danych zachodzi dla dzielnic centralnych. Należy również zauważyć, że brak odpowiadających sobie nazw przystanków w Wawrze jest spowodowany błędnie pozyskanymi danymi w BDOT10k, gdzie w nazwie przystanku nie uwzględniono jego numeru porządkowego wskazującego na kierunek jazdy danego środka transportu publicznego. Informacja taka znalazła się w zbiorze OSM z dnia 05.2015 roku powodując 100% niezgodność nazw przystanków w tej dzielnicy.

Podsumowanie i wnioski

Wzrost ilości danych OSM, jako jednego z najbardziej popularnych projektów działających na zasadzie *crowdsourcingu* jest dobrym miernikiem zaangażowania społeczeństwa w tworzenie *VGI*. Otrzymane wyniki wskazują, iż świadomość publiczna w zakresie informacji przestrzennej rośnie w szybkim tempie, jest jednak zróżnicowana ze względu na treść topograficzną i typ geometryczny pozyskiwanych danych, a także rodzaj jednostki podziału terytorialnego kraju i jej charakter. Największy wzrost liczby danych w badanym okresie zaobserwowano dla danych dotyczących miejscowości, co niestety prawdopodobnie wynika z błędnej interpretacji przez użytkowników OSM pokrycia terenu i punktów zainteresowania. Przy czym, nie towarzyszy im powtarzająca się zależność przestrzenna – dla miejscowości największy wzrost zaobserwowano na południu kraju, dla pokrycia terenu na północy, a punkty zainteresowania cieszyły się największą popularnością w miastach. Tempo wzrostu liczby danych jest większe dla obszarów miejskich. Zakładając, że największy wkład w tworzenie *VGI* na danym obszarze mają jego mieszkańcy można stwierdzić, że ich potrzeby w zakresie danych przestrzennych są zróżnicowane, ale jednocześnie szerokie. Wydaje się jednak, że użytkownicy OSM nie dostrzegają jeszcze wartości, jaką niesie część opisowa danych przestrzennych, skupiając się na tworzeniu bardzo dokładnej części geometrycznej i pozostawiając niewypełnione atrybuty obiektów lub popełniając błędy wynikające w dużej mierze z nieuwagi, co w znacznym stopniu obniża ich jakość.

Porównanie *VGI* i danych wchodzących w skład *SDI* wskazuje na wzrost podobieństwa tych zbiorów, przy czym jego tempo zależy od typu geometrycznego danych, a także charakteru gminy w przypadku danych wielkoskalowych. Największym podobieństwem charakteryzują się dane punktowe, następnie liniowe, a najmniejszym powierzchniowe. Przyczyną tego zjawiska może być fakt, że dla nieprofesjonalnych użytkowników danych przestrzennych najłatwiejsze technologicznie jest pozyskiwanie właśnie danych punktowych, a najtrudniejsze – powierzchniowych. Wzrost części wspólnej danych OSM i BDOO oraz BDOT10k świadczy o zwiększającej się jakości *VGI* w Polsce. Można również zauważyć zwiększającą się liczbę obiektów wyszczególnionych tylko w danych OSM, która wynika ze zwiększającej się różnicy aktualności porównywanych zbiorów i wskazuje na potrzebę aktualizacji zbiorów pochodzących z *SDI*. Zaprezentowana w poniższym artykule metodyka może być zastosowana do wskazywania miejsc wymagających aktualizacji w BDOO lub

BDOT10k z danych OSM. Należy ją jednak poszerzyć o dokładną analizę atrybutów TYPE i NAME, których jakość wypełnienia może stanowić poważną przeszkodę w zoptymalizowaniu tego procesu. Należy podkreślić fakt, iż mimo podobieństw OSM do zbiorów wchodzących w skład SDI dane te charakteryzują się wysoką przydatnością głównie do celów prywatnych. Nie mogą być one wykorzystane w procesach podejmowania decyzji administracyjnych niosących skutki prawne i gospodarcze, gdyż do ich podjęcia wymagane są dane o najwyższej jakości gromadzone w ramach Krajowej Infrastruktury Informacji Przestrzennej. Dane VGI mogą mieć natomiast duże znaczenie w popularyzowaniu profesjonalnych danych przestrzennych oraz uświadomieniu społeczeństwu korzyści wynikających z ich tworzenia, gromadzenia i stosowania.

Zaprezentowane w artykule podejście do oceny zaangażowania społeczeństwa w tworzenie VGI pozwala stwierdzić, iż jesteśmy świadkami wzrostu świadomości dotyczącej danych przestrzennych w Polsce. Więcej o jego poziomie i tempie można będzie wywnioskować porównując go z innymi państwami w Europie, a nawet na świecie, co może stanowić przedmiot przyszłych badań nad zjawiskiem *crowdsourcingu*. Dane VGI są obrazem potrzeb społeczeństwa w zakresie danych przestrzennych i powinny być brane pod uwagę w tworzeniu SDI. Zwłaszcza w kontekście ich wysokiej aktualności, której dane profesjonalne tworzone jedynie przez organizacje rządowe nigdy nie osiągną. Wzrost zainteresowania VGI pozwala prognozować, iż *crowdsourcing* długo będzie wymieniany wśród trendów w technologiach geoprzestrzennych. W związku z czym, zasadne wydaje się znalezienie odpowiednika określeń VGI i *crowdsourcing* w języku polskim. W kontekście zaprezentowanych badań i przeglądu literatury najbardziej odpowiedni wydaje się termin społecznościowe dane przestrzenne, którego poprawność i adekwatność autorka pozostawia ocenie środowiska.

Literatura

- Bakula K., 2014: Rola redukcji ilościowej danych wysokościowych pozyskanych z lotniczego skaningu laserowego w procesie tworzenia map zagrożenia powodziowego, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
- Castelein W., Grus L., Crompvoets J., Bregt A., 2010: A characterization of volunteered geographic information, 13th AGILE International Conference on Geographic Information Science 2010, Portugal.
- Coleman D.J., Georgiadou Y., Labonte J., 2009: Volunteered geographic information: the nature and motivation of producers, *International Journal of Spatial Data Infrastructure Research* vol. 4: 332-358.
- Craglia, M., 2007: Volunteered geographic information and spatial data infrastructures: When do parallel lines converge?, VGI Specialist Meeting, Santa Barbara, dostępnego lipiec 2015 r. http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Craglia_paper.pdf
- Craglia M., Goodchild M.F., Annoni A., Camara G., Gould M., Kuhn W., Mark D., Masser I., Maguire D., Liang S., Parsons E., 2008: Next-Generation Digital Earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science, *International Journal of Spatial Data Infrastructures Research* vol. 3: 146-167.
- Gaździcki J., 2012: Trendy rozwojowe w dziedzinie informacji geoprzestrzennej, *Roczniki Geomatyki* t. 10, z. 3, PTIP, Warszawa.
- Goodchild M.F., 2007: Citizens as sensors: the Word of volunteered geography, *GeoJournal* vol. 69.
- Goodchild M.F., Guo H., Annoni A., Bian L., Bie K., Campbell F., Craglia M., Ehlers M., Genderen J., Jackson D., Lewis A.J., Pesaresi M., Remetey-Fülöpp G., Simpson R., Skidmore A., Wang C., Woodgate P., 2012: Next-Generation Digital Earth, Proceedings of the National Academy of Sciences of the United States of America, vol. 109, No. 28: 11088-11094.
- Haklay M., 2010: How good is volunteered geographical information? A comparative study of OpenStreet-Map and Ordnance Survey datasets, *Environment and Planning B: Planning and Design* vol. 37: 682-703.

- Ho S., Rajabifard A., 2010: Learning from the crowd: the role of volunteered geographic information in realising a spatially enabled society, GSDI 12 World Conference-Realizing spatially enabled societies: 1-23.
- Kukułka M., Gotlib D., 2014: Wpływ zjawiska neokartografii na rozwój serwisów internetowych udostępniających informacje przestrzenne, *Polski Przegląd Kartograficzny* t. 46, nr 1: 34-46.
- 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.

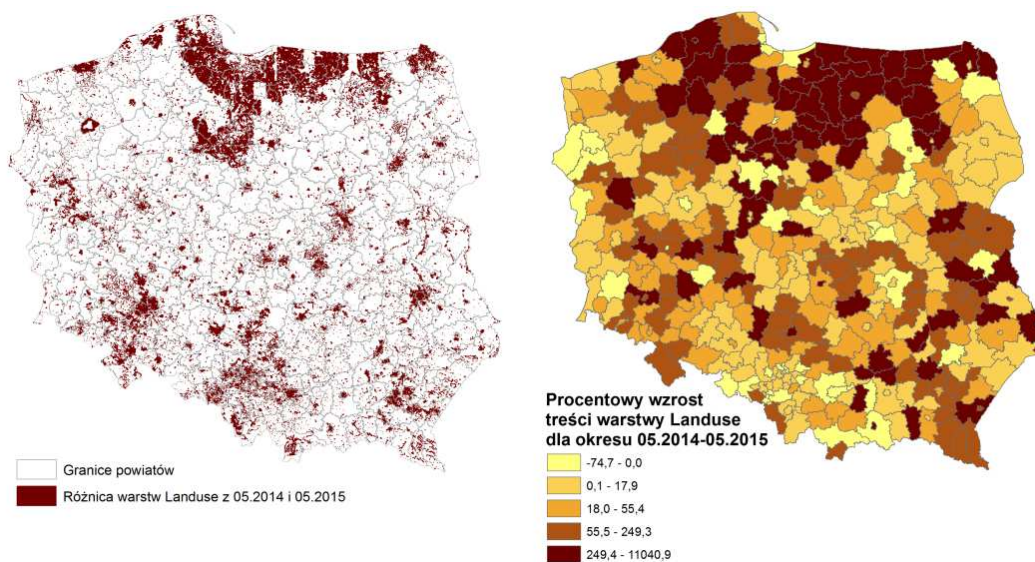
Streszczenie

W artykule przedstawiono zjawisko pozyskiwania danych przestrzennych przez osoby niebędące profesjonalistami na przykładzie projektu OpenStreetMap. Dzięki technologii Web 2.0 głównymi twórcami tego przestrzennego serwisu internetowego są jego użytkownicy. Obserwowany w ostatnich latach wzrost zainteresowania społeczeństwa informacją geograficzną przyczynił się do powstania terminu „volunteered geographic information” (VGI), przez który należy rozumieć dane przestrzenne pozyskane przez użytkowników Internetu. Zjawisko, opisywane tylko w literaturze zagranicznej, nabiera coraz większego znaczenia również w Polsce. W artykule poddano ocenie wzrost treści i jakości danych przestrzennych z projektu OpenStreetMap w latach 2014-2015, a także porównano szczegółowość oraz liczbę pozyskiwanych danych w odniesieniu do jednostek podziału terytorialnego kraju. Na podstawie otrzymanych wyników, wnioskowano o zaangażowaniu w tworzenie danych przestrzennych, użytkowników z różnych regionów Polski. Ponadto dane z OpenStreetMap porównano ze zbiorami danych przestrzennych wchodzącymi w skład Zasobu Geodezyjnego i Kartograficznego – BDOT10k i BDOO.

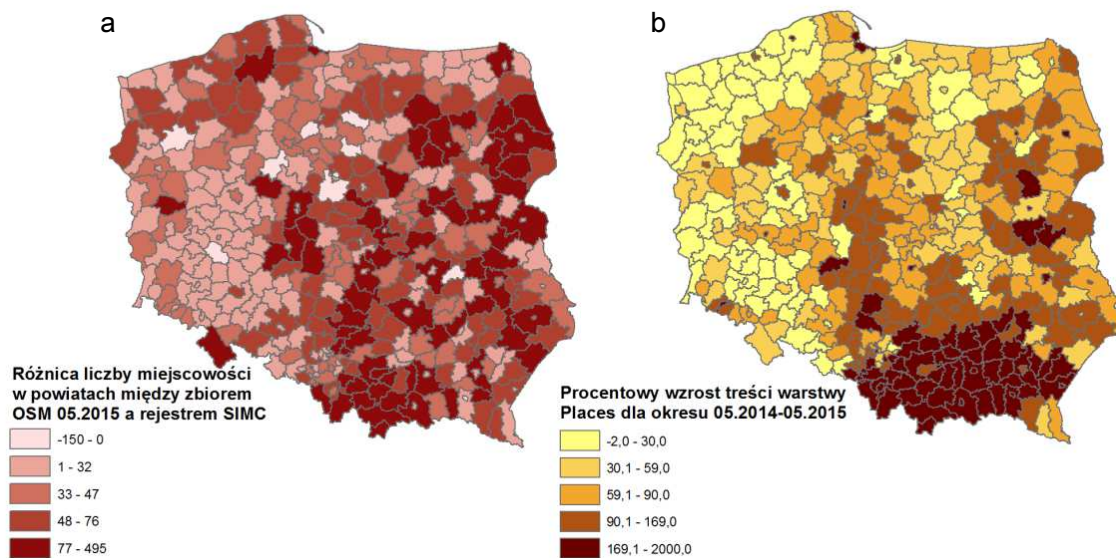
Abstract

The article presents the phenomenon of spatial data acquisition by non-professionals using the example of the OpenStreetMap project. Thanks to Web 2.0 technology the main creators of this web service are its users. An increased interest in geographic information has been observed recently. It contributed to the creation of the term „volunteered geographic information” (VGI), which should be understood as the spatial data acquired by Internet users. The phenomenon, described only in foreign literature, is becoming increasingly important also in Poland. The article assesses the growth of content and quality of spatial data from the OpenStreetMap project in the period 2014-2015, and compares the level of detail and the amount of data obtained in relation to territorial division units of the country. On the basis of the results commitment of Polish society to the creation of spatial data in different parts of Poland was concluded. In addition, OpenStreetMap data was compared with BDOT10k and BDOO, the spatial datasets included in the Polish spatial data infrastructure.

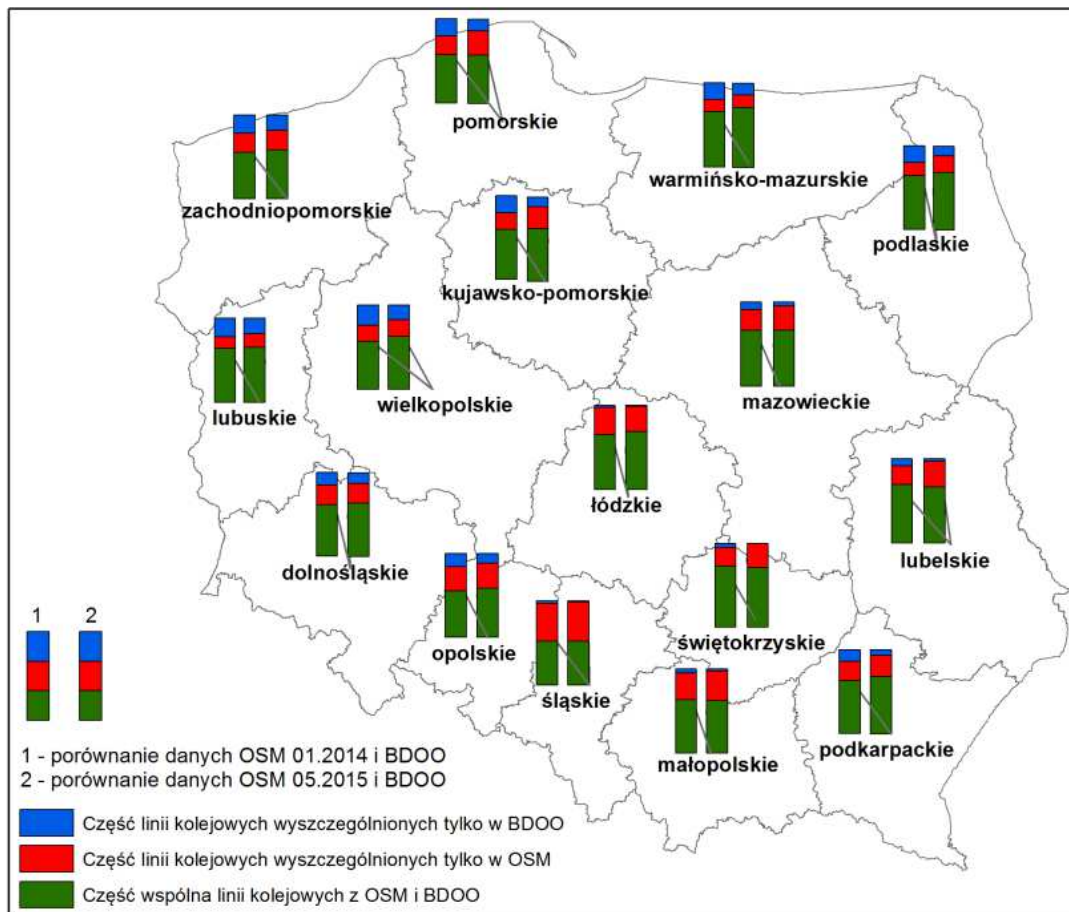
mgr inż. Sylwia Marczak
s.marczak@gik.pw.edu.pl



Rysunek 3. Wzrost ilości danych warstwy *Landuse* w okresie 05.2014–05.2015 roku



Rysunek 4. a – porównanie liczby miejscowości w OSM i SIMC, b – wzrost ilości danych warstwy *Placess* w okresie 05.2014–05.2015 roku



Rysunek 5. Porównanie treści klasy obiektów SKTR_L z BDOO i warstwy *Railways* z OSM