

Porównanie zaangażowania obywateli wybranych krajów europejskich w tworzenie VGI

Comparison of public involvement of selected european countries in VGI creation

Sylwia Marczak

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

Słowa kluczowe: OpenStreetMap, społecznościowe dane przestrzenne, analiza aktywnych użytkowników OSM

Keywords: OpenStreetMap, volunteered geographic information, analysis of OSM contributors

Wstęp

Wraz z początkiem XXI wieku nastąpił przełom technologiczny, który przyczynił się do późniejszego, nadal trwającego, upowszechniania wiedzy geoprzestrzennej wśród społeczeństwa. Było to możliwe dzięki udostępnieniu cywilom niezdegradowanego sygnału z systemu GPS (*Global Positioning System*) w 2000 roku. Wywołało to szybki rozwój urządzeń pozycjonujących, z których użytkownicy mogą korzystać, każdy w dowolnym celu. Oprócz zastosowań oczywistych jak: nawigacja, kartografia lub geodezja, sygnał został wykorzystany również w celach rekreacyjnych, takich jak *geocaching*, który polega na poszukiwaniu tzw. skrytek uprzednio ukrytych przez innych uczestników zabawy, których lokalizacja została udostępniona za pomocą współrzędnych geograficznych. Z czasem każdy telefon komórkowy został wyposażony w odbiornik sygnału GPS, co umożliwiło przypisywanie informacji o położeniu do zdjęć (tzw. geotagowanie) lub innych informacji (Neis, Zielstra, 2014). Drugim istotnym czynnikiem mającym wpływ na upowszechnienie wiedzy geoprzestrzennej była technologia Web 2.0. Terminem tym określa się potocznie serwisy internetowe powstałe po 2001 roku. Umożliwiła ona użytkownikom Internetu nie tylko korzystanie z zawartych w nim informacji, ale również tworzenie nowych lub edytowanie istniejących (Neis, Zielstra, 2014). Znacząco przyczyniło się to do rozwoju serwisów społecznościowych, z których najpopularniejsze to oczywiście Facebook i Twitter, ale również YouTube i Wikipedia, a także OpenStreetMap (OSM), który daje użytkownikom możliwość tworzenia danych przestrzennych.

Dotychczas powstało kilka terminów definiujących dane zbierane przez użytkowników Internetu. W literaturze światowej używane są określenia *user-generated content* (Anderson, 2007) lub *user-created content* (Wunsch-Vincent, Vickery, 2007). Do określenia danych przestrzennych pozyskiwanych w ten sposób używane są terminy *volunteered geographic information* (VGI, Goodchild, 2007) lub *crowd-sourced geodata* (Hudson-Smith i in., 2008). Polskim odpowiednikiem tych terminów może być „społecznościowe dane przestrzenne” (Marczak, 2015).

Zjawisko zbierania danych przestrzennych przez amatorów – nieposiadających wiedzy profesjonalnej (z zakresu ogólnie rozumianej geomatyki) – użytkowników Internetu, leży od kilku lat w kręgu zainteresowań naukowców z całego świata. Zdecydowanie największa liczba badań dotyczy najpopularniejszego serwisu, umożliwiającego społeczeństwu zbieranie danych przestrzennych jakim jest OpenStreetMap. Należy przy tym zauważyć, że tematyka ta poruszana jest głównie w literaturze zagranicznej i dotyczy różnych aspektów związanych z VGI, takich jak: analiza jakości danych, analiza zaangażowania użytkowników oraz różnego rodzaju możliwości zastosowań społecznościowych danych przestrzennych (Neis, Zielstra, 2014). Ocenę jakości danych VGI przeprowadzano w porównaniu do danych urzędowych (m.in. Haklay, 2010; Girres, Touya, 2010; Marczak, 2015; Nowak Da Costa i in., 2016) lub komercyjnych (m.in. Zielstra, Zipf, 2010; Ludwig i in., 2011; Neis i in. 2012). Z przytoczonych prac dotyczących Wielkiej Brytanii, Niemiec, Francji i Polski, można wysnuć jeden wspólny wniosek – dane OSM są wysoko heterogeniczne – obszary miejskie charakteryzują się danymi o wysokiej jakości i dużym pokryciu, w przeciwieństwie do danych dotyczących obszarów wiejskich. Natomiast jak podaje Arsanjani i in. (2015) można jednoznacznie stwierdzić, że projekt OpenStreetMap zrewolucjonizował i poszerzył sposoby zbierania danych przestrzennych. Badania dotyczące użytkowników OSM skupiały się na aspekcie nierównomierności zaangażowania (m.in. Neis, Zipf, 2012; Budhathoki, 2010), rozkładzie liczby użytkowników na świecie (Arsanjani i in., 2015; Neis, Zipf, 2012; Budhathoki, 2010) oraz ich motywacjach i zachowaniach (m.in. Steinmann i in., 2013; Schmidt, Klettner, 2013; Lin, 2011; Schmidt i in., 2013). Wspólnym mianownikiem wszystkich tych rozważań jest stwierdzenie, że tak jak w przypadku danych, również społeczność OSM jest wysoko heterogeniczna zarówno pod względem liczby użytkowników na świecie, jak i motywacji lub poziomu zaangażowania. Różnego rodzaju zastosowaniom danych OSM również poświęcono wiele pozycji w literaturze, badania te dotyczyły różnych dziedzin począwszy od wykorzystania OSM w tworzeniu danych o użytkowaniu terenu (Arsanjani, 2014), przez analizy możliwości aplikacji danych OSM w analizach sieciowych (Cichociński, 2012; Cichociński, Dębińska, 2012) i budowaniu modeli miast 3D (Schilling i in., 2009). W literaturze, do której dotarła autorka tylko kilka pozycji dotyczyło porównań międzynarodowych danych i użytkowników OSM. Neis i in. (2013) porównywali rozwój OSM od początku jego istnienia do października 2012 roku w 12 miastach świata, zauważając w miastach europejskich dużą koncentrację danych, które tworzone są głównie przez użytkowników mieszkających na terenie tych miast lub w bezpośrednim sąsiedztwie. Z kolei Arsanjani i Vaz (2015) porównywali kompletność danych o użytkowaniu terenu również dla obszarów miejskich, ale położonych tylko na terenie Europy. Natomiast Yang i in. (2016) dokonali analizy czasowej i porównania nierównomierności zaangażowania użytkowników OpenStreetMap dla czterech krajów stwierdzając, że ma ona tendencję wzrostową dla tych z nich, w których nie importowano do bazy OSM danych z innych źródeł (Niemcy), a wydaje się być losowa dla krajów, w których odsetek danych importowanych jest wysoki (Stany Zjednoczone i Holan-

dia). Celem niniejszego artykułu jest uzupełnienie tych badań o paneuropejską analizę czasową zaangażowania użytkowników OSM w tworzenie VGI dla dziesięciu regionów Europy na poziomie NUTS 1 wg Klasyfikacji Jednostek Terytorialnych do Celów Statystycznych. W zależności od kraju za region NUTS 1 przyjęto jednostki administracyjne (np. niemieckie landy) bądź połączenie kilku jednostek administracyjnych (np. grupy województw w Polsce).

Zaangażowanie użytkowników OSM będzie mierzone trzema parametrami – liczbą aktywnych użytkowników, przyrostem danych przestrzennych przez nich tworzonych oraz ich efektywnością. W związku ze stale rosnącym zainteresowaniem społeczności internetowej projektem OSM należy spodziewać się wzrostu zaangażowania użytkowników Internetu w tworzenie danych przestrzennych.

Wybór obszarów badań

Zanim wybrano obszary badań założono, że im wyższy poziom zaawansowania ludności w korzystaniu z komputera i Internetu, tym większa liczba tworzonych danych i liczba użytkowników OSM. W związku z tym, wybór obszarów badań poprzedzono analizą poziomu zaawansowania w korzystaniu z komputera i Internetu przez społeczeństwa krajów Europy. Miało to za zadanie wyeliminować wpływ obicia technologicznego mieszkańców na zaangażowanie w tworzenie społecznościowych danych przestrzennych. Drugim oczywistym czynnikiem mającym wpływ na wielkość kontrybucji użytkowników OSM jest liczba ludności, jednak jej wpływ można łatwo wyeliminować używając danych względnych do niej odniesionych.

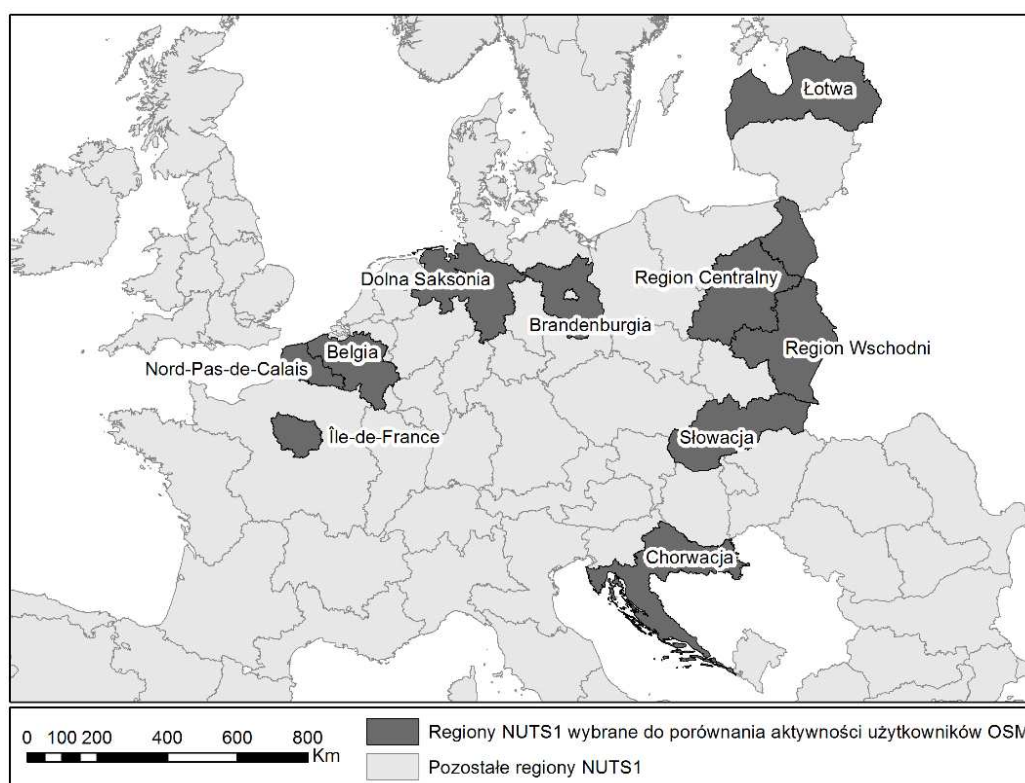
Analizę poziomu zaawansowania technologicznego ludności przeprowadzono w dwóch etapach, najpierw wybrano sześć krajów Europy najbardziej podobnych do Polski pod względem korzystania z komputera i Internetu, a następnie w tych krajach, w których było to możliwe wybrano po dwa regiony NUTS 1, charakteryzujące się najniższym i najwyższym procentem gospodarstw domowych mających dostęp do Internetu. W obu etapach skorzystano z danych statystycznych pozyskanych z Urzędu Statystycznego Unii Europejskiej – Eurostatu. Pomimo tematycznych ograniczeń w dostępności danych, w pierwszym etapie wzięto pod uwagę następujące cechy społeczeństwa:

- procent społeczeństwa w wieku 16-29 lat codziennie używający komputera/Internetu,
- procent społeczeństwa w wieku 16-29 lat, który stworzył (przynajmniej raz) stronę internetową,
- procent społeczeństwa w wieku 16-29 lat, który korzysta z wyszukiwarek internetowych w celu pozyskania informacji,
- procent społeczeństwa w wieku 16-29 lat, który posiada co najmniej 3 umiejętności korzystania z Internetu (analizowane umiejętności: tworzenie stron WWW, korzystanie z wyszukiwarek, wysyłanie maili z załącznikiem, dodawanie postów na grupach dyskusyjnych, czatach, dzwonienie przez Internet),
- procent społeczeństwa w wieku 16-29 lat, który napisał program komputerowy przy użyciu języka programowania,
- procent społeczeństwa w wieku 16-29 lat, który umieszcza na stronach internetowych stworzone przez siebie treści,

- procent społeczeństwa w wieku 25-54 lata z wykształceniem podstawowym/średnim/wyższym, który tworzy strony WWW lub blogi,
- procent osób, używających Internetu w ciągu ostatnich 3 miesięcy (rok 2015), które korzystały z chmury do przechowywania dokumentów, muzyki, filmów.

Dane roczne (z lata 2011-2015) uśredniono w ramach jednej kategorii dla wszystkich dostępnych lat, a następnie policzono bezwzględne różnice między wartościami statystyk dla Polski i krajów europejskich. Ostatecznie uzyskane różnice zsumowano i zidentyfikowano pięć krajów najbardziej podobnych do Polski pod względem zaawansowania ludności w korzystaniu z komputera i Internetu. Były to w kolejności malejącego podobieństwa: Słowacja, Chorwacja, Belgia, Francja, Niemcy i Łotwa.

Kolejnym krokiem był wybór regionów NUTS 1 w poszczególnych krajach. Dla trzech z nich – Słowacji, Chorwacji i Łotwy – poziom regionalny NUTS 1 jest równy poziomowi NUTS 0 obejmującemu cały kraj (rys. 1). Ponadto w Belgii na poziomie NUTS 1 zostały wyodrębnione tylko 3 jednostki, w tym jedną z nich był obszar aglomeracji Brukseli, co skłoniło autorkę do dalszych analiz na poziomie krajowym w przypadku tego państwa (rys. 1). Natomiast dla Polski, Niemiec i Francji wybrano regiony NUTS 1 o najniższych i najwyższych wartościach następujących cech:



Rysunek 1. Regiony NUTS 1 wybrane do analizy zaangażowania obywateli w tworzenie społecznościowych danych przestrzennych (źródło: opracowanie własne na podstawie danych Eurostat)

- procent gospodarstw domowych z szerokopasmowym dostępem do Internetu,
- procent gospodarstw domowych z dostępem do Internetu,
- procent społeczeństwa korzystający z Internetu co najmniej raz w tygodniu.

Wybór cech był podyktowany dostępnością danych statystycznych na poziomie regionalnym i ich potencjalnym wpływem na tworzenie społecznościowych danych przestrzennych. Regiony o najniższych wartościach powyższych cech to (rys. 1): Region Wschodni (złożony z województw: podkarpackiego, świętokrzyskiego, lubelskiego i podlaskiego), Brandenburgia (niemiecki land) i Nord-Pas-de-Calais (jedna z jednostek najwyższego stopnia podziału terytorialnego we Francji). Regiony o najwyższych wartościach powyższych cech to (rys. 1): Region Centralny (złożony z województw: mazowieckiego i łódzkiego), Dolna Saksonia (Niemcy) i Île de France (Francja).

Dane z projektu OpenStreetMap wykorzystane do analizy

Projekt OpenStreetMap jest wyczerpująco opisany zarówno w literaturze zagranicznej (m.in. Neis, Zielstra, 2014; Neis i in. 2013) jak i polskiej (m.in. Cichociński, 2012; Nowak Da Costa i in., 2016), dlatego w tym artykule zostaną przedstawione tylko najważniejsze informacje. Serwis OpenStreetMap działa od 2004 roku i jest obecnie najpopularniejszym projektem działającym w oparciu o crowdsourcing na świecie. 17 sierpnia 2016 roku miał 2 961 615 zarejestrowanych na całym świecie użytkowników, którzy stworzyli niemal 3,5 miliarda węzłów (*nodes*) i ponad 3,5 miliona linii. Na strukturę bazy danych OSM składają się trzy rodzaje obiektów: węzły (*node*), linie (*way*) i relacje (*relation*). Do tworzenia obiektów punktowych służą węzły, natomiast liniowych i poligonowych – linie. Relacje służą do określania powiązań między obiektami. Informacja o cechach obiektów zapisywana jest w postaci tzw. tagów, przyjmujących postać par klucz-wartość (Cichociński, 2012). Na przykład para `railway="tram"` oznacza obiekty będące torami tramwajowymi. W celu standaryzacji danych, użytkownikom zaleca się korzystanie z dostępnych par klucz-wartość do opisu nowotworzonych obiektów, ale istnieje możliwość zaproponowania własnych (Neis, Zielstra, 2014).

Opisany powyżej model danych OSM znacznie różni się od modelu relacyjnego powszechnie stosowanego w systemach informacji geograficznej, co sprawia, że korzystanie z danych w oprogramowaniu GIS stwarza pewne trudności. Istnieje jednak możliwość skorzystania z danych przetworzonych do formatu .shp darmowo udostępnianych przez firmę Geofabrik, niestety ich zakres jest ograniczony zaledwie do 8 klas obiektów. Drugi dostępny format danych to .osm xml, w którym pobrano dane dla wybranych obszarów badań o aktualności na dzień 19.07.2016 r. W celu przeprowadzenia analizy dane zaimportowano do geobazy plikowej korzystając z zestawu narzędzi programu ArcGIS – „ArcGIS Editor for OSM”, stworzonego i darmowo udostępnianego przez firmę Esri. Jego narzędzia pozwalają pobierać, edytować i automatycznie nadawać symbolizację danym OSM. Należy przy tym zwrócić uwagę na fakt, że duża objętość danych (od 1 do 6 GB zależnie od obszaru) znacząco utrudniała i wydłużała pracę z nimi w oprogramowaniu GIS, co skłania autorkę do wniosku, iż w celu analizy danych OSM należy korzystać z innych rozwiązań, na przykład oprogramowania bazodanowego. Zaimportowane za pomocą narzędzia „Load OSM File” dane, zapisywane były w geobazie plikowej w trzech klasach obiektów – punktowych, liniowych i poligonowych oraz tabeli, w której zapisana była informacja o relacjach łączących obiekty.

Informacje opisowe zapisane były w każdej klasie obiektów w identyczny sposób: jako kolumny (atrybuty) zapisane zostały „klucze” z pary klucz-wartość przypisane do obiektów, natomiast część „wartość” przypisano wartościom atrybutów. Dodatkowe informacje zawarte w tabeli atrybutów to identyfikator obiektu, nazwa użytkownika tworzącego obiekt oraz data i godzina utworzenia obiektu. Dzięki temu ostatniemu atrybutowi możliwe było przeprowadzenie analizy czasowej, która zakładała porównanie informacji o ilości danych i liczbie użytkowników w siedmiu półrocznych okresach, począwszy od pierwszej połowy (do 30 czerwca włącznie) 2013 roku, a skończywszy na pierwszej połowie roku 2016.

Do analizy wybrano obiekty punktowe, liniowe i poligonowe o kluczach *natural* – obiekty naturalne, *railway* – obiekty związane z transportem kolejowym i *historic* – obiekty historyczne. Pomimo faktu, że obiektom tym można przyporządkować jeden główny typ geometryczny (obiekty naturalne – poligonowe, obiekty związane z transportem kolejowym – liniowe, a historyczne – punktowe), dla każdego klucza istniały dane o wszystkich typach geometrycznych. Sprawilo to, że dla każdego klucza oddzielnie analizowano obiekty punktowe, liniowe i poligonowe, w sumie było to zatem 9 klas obiektów.

Porównanie liczby aktywnych użytkowników OSM w Europie

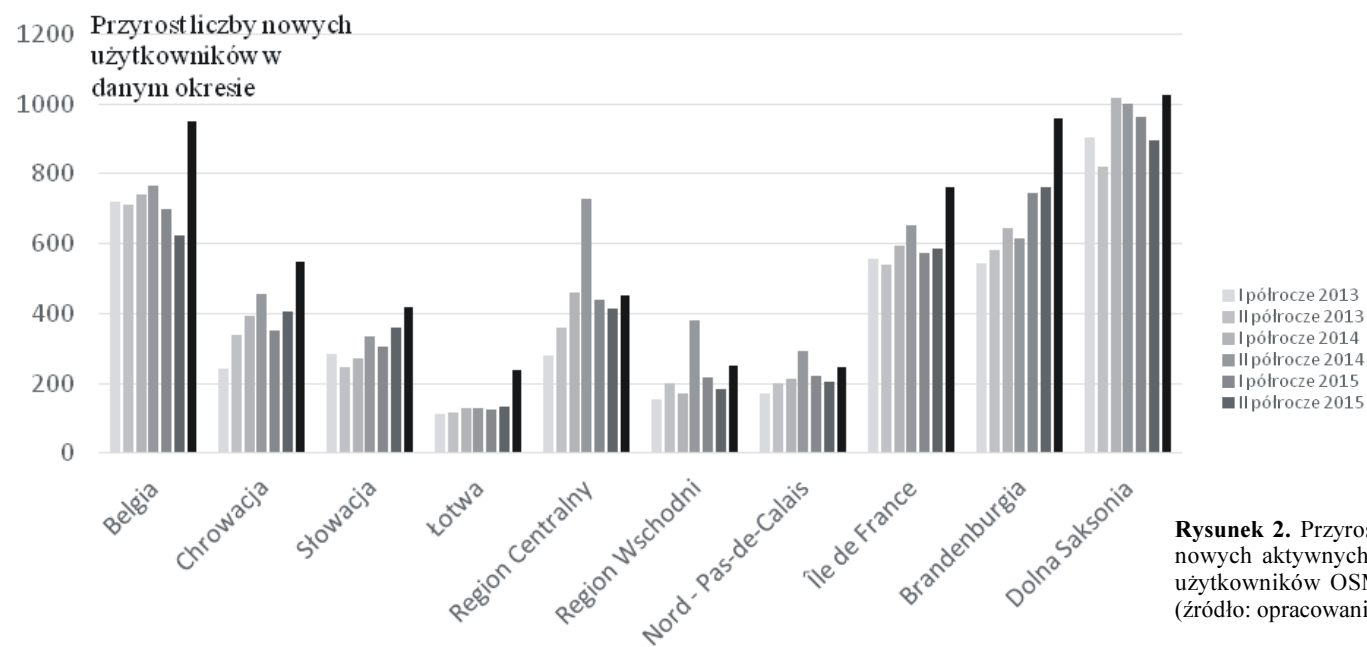
Dotychczasowe analizy opisane w literaturze dotyczące liczby użytkowników OpenStreetMap przeprowadzano dzieląc grupę wszystkich zarejestrowanych użytkowników na podgrupy (Neis, Zipf, 2012; Neis, Zielstra, 2014; Yang i in., 2015). Niestety, nie wypracowano jednych ogólnie obowiązujących kryteriów podziału. W artykule skorzystano z podziału zaproponowanego w pracy Neis i Zielstra (2014), która wyróżnia trzy grupy użytkowników: 1) użytkownicy danych (*users*) – korzystają z danych VGI, ale ich nie edytują; 2) aktywni użytkownicy (*contributors*) – posiadają konto w serwisie VGI, edytują i tworzą dane; 3) zarejestrowani w serwisie (*registered members*) – posiadają konto w serwisie VGI, ale nie edytują i nie tworzą danych. Zbadano jak zmieniała się liczba aktywnych użytkowników – takich którzy stworzyli lub edytowali przynajmniej jeden obiekt w danych OSM w wybranych obszarach w czasie. Należy zwrócić uwagę na fakt, że baza danych OSM nie zbiera informacji o miejscu zamieszkania użytkowników. Oznacza to, że stworzenie obiektu w danym kraju nie jest równoznaczne z tym, że użytkownik, który go stworzył pochodzi z tego kraju. W literaturze opisywane są algorytmy wyznaczania pochodzenia użytkowników na podstawie danych przez nich tworzonych (m.in. Neis, Zipf, 2012; Neis i in., 2013). Wyniki tych badań wskazują, że w Europie dane OSM tworzone są przez lokalnych użytkowników (*local mappers*), a więc mieszkańców danego kraju bądź regionu (Neis i in., 2013). W związku z tym w artykule przyjęto założenie, że liczba aktywnych użytkowników OSM świadczy o zaangażowaniu społeczeństwa danego regionu w tworzenie danych VGI.

W celu przeprowadzenia analizy, zsumowano liczbę nowych aktywnych użytkowników w danych OSM pobranych dla każdego regionu, w podziale na 7 wcześniej zdefiniowanych okresów. Za nowego użytkownika uznawano tego, który w okresie poprzedzającym nie wprowadził żadnych zmian w danych OSM. Obliczeń dokonano za pomocą narzędzia *Summary Statistics* z pakietu ArcGIS.

Do końca pierwszej połowy 2016 roku największa liczba aktywnych użytkowników występowała w Dolnej Saksonii (13 245) i Belgii (8910), natomiast najmniejsza na Łotwie

Tabela 1. Statystyki dotyczące liczby użytkowników OSM w regionach wybranych do analizy (źródło: opracowanie własne)

Kraj/Region	Belgia	Chorwacja	Słowacja	Łotwa	Region Centralny	Region Wschodni	Nord-Pas-deCalais	Île de France	Brandenburgia	Dolna Saksonia
Liczba aktywnych użytkowników	8910	4257	3464	1642	4360	2117	2603	7209	8555	13245
Liczba aktywnych użytkowników/1 mln ludności	791	1007	639	827	559	320	639	597	3481	1692
Średni półroczny przyrost liczby nowych użytkowników OSM w okresie od 01.2013 do 06.2016 r.	744	391	318	141	448	224	222	610	693	947



Rysunek 2. Przyrost liczby nowych aktywnych użytkowników OSM (źródło: opracowanie własne)

(1642) i w Regionie Wschodnim (2117). Po uwzględnieniu liczby ludności danego regionu, stwierdzono że największa liczba aktywnych użytkowników OSM na 1 milion ludności występuje w Brandenburgii (3481) i Dolnej Saksonii (1692), natomiast najmniejsza wynosząca zaledwie 320 w Regionie Wschodnim i 559 w Regionie Centralnym (tab. 1).

W związku ze stale rosnącą popularnością projektu OSM oczywiste jest, że liczba nowych aktywnych użytkowników we wszystkich regionach wzrastała, natomiast przyrosty te są silnie zróżnicowane zarówno geograficznie, jak i czasowo (rys. 2). Zakres przyrostów liczby nowych aktywnych użytkowników OSM wynosił od 112 do 1026, przy czym dla każdego regionu tempo wzrostu było inne. Średni półroczny przyrost nowych aktywnych użytkowników OSM wykazuje duże zróżnicowanie, z wartością minimalną dla Łotwy i maksymalną dla Dolnej Saksonii (tab. 1).

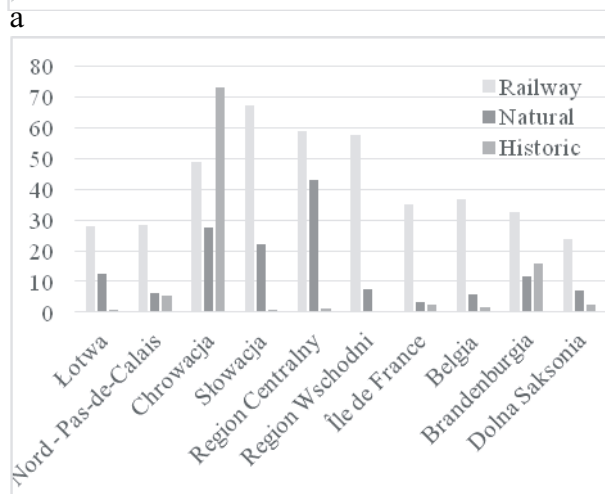
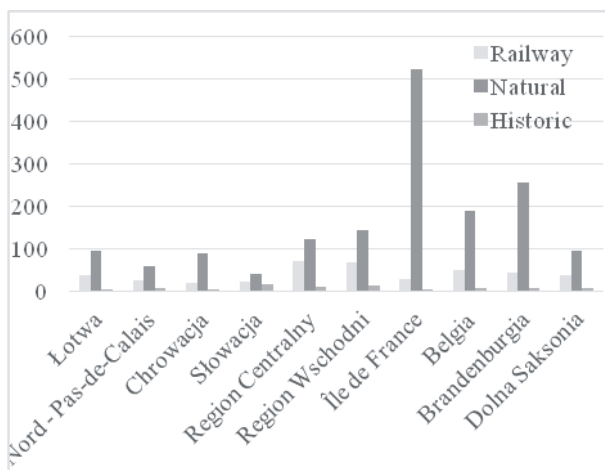
Dla wszystkich regionów oprócz Centralnego, Wschodniego i Nord-Pas-de-Calais w pierwszej połowie 2016 roku zaobserwowano najwyższy przyrost liczby nowych użytkowników OSM. Natomiast dla trzech wyżej wymienionych regionów była to druga połowa 2014 roku. W odniesieniu do najniższego przyrostu liczby nowych użytkowników nie można zaobserwować równie systematycznej zależności.

Ocena efektywności użytkowników OSM

Do oceny efektywności użytkowników OSM przyjęto następującą miarę – średnia liczba/długość/powierzchnia obiektów stworzona przez jednego aktywnego użytkownika analizowanego rodzaju obiektów. Również w tej analizie przyjęto założenie, że dane OSM tworzą mieszkańcy danego regionu. Jednak należy również zwrócić uwagę na fakt, że praca wolontariuszy nie jest jedynym sposobem pozyskiwania danych do bazy OSM. Drugim sposobem jest ich import z innych baz danych, urzędowych lub komercyjnych, co zgodnie z literaturą miało miejsce dla niektórych z badanych regionów lub ich części (Neis i in, 2012; Cichociński, 2012). Oznacza to, że wykonana ocena efektywności jest w pewnym sensie „zanieczyszczona” wpływem importowanych danych, ale mimo wszystko pozwala wykryć pewne zależności i tendencje.

Analizy dokonano oddzielnie dla danych punktowych, liniowych i poligonowych o wcześniej wybranych tagach – *natural*, *railway* i *historic*. W tym celu z zaimportowanych danych OSM wyselekcjonowano te o pożądanym tagach, a następnie obliczono: liczbę, długość lub powierzchnię obiektów w zależności od typu geometrycznego oraz liczbę użytkowników je tworzących. Otrzymane wyniki nie pozwalają wskazać regionu o najbardziej efektywnych użytkownikach, gdyż występują znaczne różnice w ramach jednego kraju w zależności od typu geometrycznego danych i tagów. Dla danych punktowych największa efektywność, ale również największe zróżnicowanie pomiędzy krajami występuje dla obiektów naturalnych, których średnio najwięcej tworzy użytkownik regionu Île de France (rys. 3a). W przypadku danych liniowych największa efektywność wśród użytkowników badanych regionów występuje dla obiektów związanych z transportem kolejowym, których najwięcej tworzą użytkownicy w Słowacji i dwóch regionach Polski (rys. 3b).

Należy również zwrócić uwagę na wyjątkowo wysoką aktywność użytkowników Chorwacji w tworzeniu liniowych obiektów historycznych (rys. 3b). Niestety po przeanalizowaniu danych OSM użytych do analizy okazało się, że wprowadzone obiekty zostały oznaczone błędnymi tagami – do obiektów historycznych przypisano drogi, co przełożyło się na otrzymany wynik.



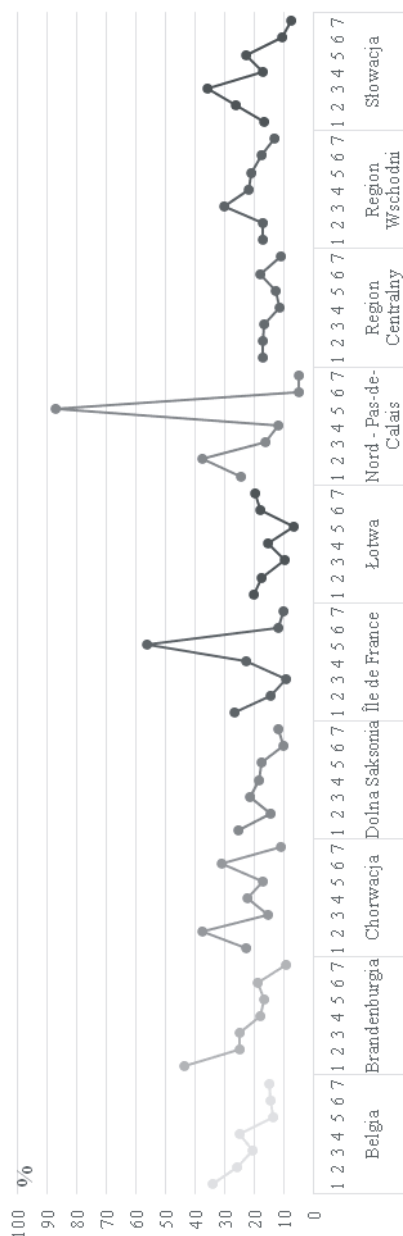
Rysunek 3. Efektywność użytkowników OSM:
 a – średnia liczba obiektów punktowych stworzonych przez 1 użytkownika,
 b – średnia długość obiektów liniowych stworzonych przez 1 użytkownika
 (źródło: opracowanie własne)

W przypadku danych poligonowych we wszystkich analizowanych regionach efektywność użytkowników w zakresie tworzenia obiektów historycznych i kolejowych jest znikoma, co spowodowane jest małą liczbą tego rodzaju obiektów o poligonowej reprezentacji geometrycznej. Odwrotna sytuacja występuje dla obiektów naturalnych, w których dominującym typem geometrycznym jest poligon. Średnia największa powierzchnia obiektów naturalnych jest tworzona przez jednego użytkownika ze Słowacji i wynosi 1743 ha, natomiast najmniejsza przez użytkownika z regionu Île de France (31,5 ha). Jest to najprawdopodobniej związane z miejskim charakterem tego regionu, gdzie większość obiektów naturalnych reprezentowanych jest w bazie danych OSM przez punkty, na co wskazała analiza dla danych punktowych.

Analiza przyrostu danych OSM w czasie

W ostatniej analizie założono, że o zainteresowaniu i zaangażowaniu użytkowników OSM w tworzenie danych przestrzennych może świadczyć przyrost treści przez nich tworzonych. W związku z tym wykonano analizę procentowego przyrostu danych OSM o wybranych

nymi tagach we wcześniej zdefiniowanych okresach. W tym celu dane podzielono według daty ich utworzenia odczytanej z atrybutu *timestamp* i przyporządkowano do jednego z półrocznych okresów od początku 2013 roku do końca pierwszej połowy 2016 roku. Następnie zsumowano liczby obiektów, długości linii i powierzchnię danych poligonowych i obliczono procentowe przyrosty danych w podziale na obiekty o tagach *historic*, *natural* i *railway*. Dla obiektów punktowych o kluczu *railway*, dla wszystkich regionów oprócz Nord-Pas-de-Calais, zaobserwowano, że z półrocza na półrocze tempo przyrostu danych rosło. Odwrotna sytuacja wystąpiła dla obiektów punktowych o kluczu *natural*, gdzie tylko w dwóch regio-



Rysunek 4. Przyrost punktowych danych OSM o kluczu *historic* w okresie I półrocze 2013-I półrocze 2016 roku (źródło: opracowanie własne)

ło 30%. W pierwszej połowie roku 2014 nastąpił bardzo znaczący wzrost liczby danych w Brandenburgii – 214% i Słowacji 139%, podczas gdy dla większości pozostałych krajów oscylował wokół 50%, co również było wartością powyżej średniej. Po sprawdzeniu liczby użytkowników okazało się, że te gwałtowne przyrosty nie były związane z importem danych z innych baz, zatem wynikały z większego zaangażowania użytkowników w tym okresie, praktycznie we wszystkich badanych regionach.

nach – Brandenburgii i Łotwie – zaobserwowano ogólny wzrost tempa przyrostu danych (tab. 2). Dla obiektów o kluczu *historic*, mimo okresowych wzrostów tempa przyrostu danych, trend dla wszystkich regionów był ujemny (rys. 4, tab. 2).

W przypadku obiektów liniowych związanych z transportem kolejowym, zaobserwowano wzrosty ilości danych między sąsiednimi okresami niejednokrotnie przekraczające 100%. Dla Regionu Centralnego było to aż 1310% w I połowie 2014 roku, a dla Wschodniego 420% w I połowie 2015 roku. Zmiany te w przypadku regionu Centralnego zostały wprowadzone przez 19 użytkowników, a w przypadku regionu Wschodniego przez 41 użytkowników. W obydwu przypadkach czas tworzenia danych trwał kilkanaście dni na przestrzeni całego półrocza. Można z tego zatem wywnioskować, że zmiana ta nie wynikała z importu danych z baz danych zewnętrznych, a z wyjątkowego wzrostu aktywności użytkowników OSM. Trend tempa przyrostu danych liniowych o kluczu *railway* w badanym okresie jest dodatni dla Belgii, Brandenburgii, Słowacji i Regionu Wschodniego, a ujemny dla pozostałych obszarów (tab. 2). Dla obiektów liniowych o tagu *natural* średnio największe przyrosty danych wystąpiły w dwóch pierwszych okresach, tj. w roku 2013. Natomiast najmniejszy średni wzrost liczby danych wystąpił w pierwszej połowie 2016 roku i wyniósł 31%. Jeśli chodzi o trendy przyrostu danych, to są one dodatnie dla Chorwacji, Brandenburgii i Regionu Centralnego i ujemne dla pozostałych obszarów (tab. 2).

W przypadku obiektów naturalnych, poligonowych, można stwierdzić, iż wzrost ich powierzchni w czasie był wyrównany i średnio dla wszystkich regionów i we wszystkich okresach, oprócz I połowy 2014 roku, wynosił oko-

Tabela 2. Trend przyrostu danych w okresie 01.2013-06.2016 r. \triangle – trend dodatni, \blacktriangledown – trend ujemny (źródło: opracowanie własne)

Typ danych	Punktowe			Liniowe*		Poligonowe		
	Tag							
Kraj	<i>Railway</i>	<i>Natural</i>	<i>Historic</i>	<i>Railway</i>	<i>Natural</i>	<i>Railway</i>	<i>Natural</i>	<i>Historic</i>
Belgia	\triangle	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\triangle	\blacktriangledown	\blacktriangledown
Chorwacja	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown
Łotwa	\triangle	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\blacktriangledown
Słowacja	\triangle	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\triangle	\triangle	\triangle
Nord-Pas-de Calais	\blacktriangledown	\blacktriangledown	\blacktriangledown	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown
Île de France	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\triangle
Region Centralny	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\triangle	\triangle
Region Wschodni	\triangle	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle
Brandenburgia	\triangle	\triangle	\blacktriangledown	\triangle	\triangle	\blacktriangledown	\blacktriangledown	\triangle
Dolna Saksonia	\triangle	\blacktriangledown	\blacktriangledown	\blacktriangledown	\blacktriangledown	\triangle	\blacktriangledown	\blacktriangledown

* Zbyt mała liczba danych dla historycznych obiektów liniowych uniemożliwiła wykonanie obliczeń.

Podsumowanie i wnioski

W artykule podjęto próbę oceny zaangażowania użytkowników OSM w tworzenie społecznościowych danych przestrzennych w różnych regionach Europy, w okresie od stycznia 2013 do końca czerwca 2016 roku. Dokonano tego analizując wzrost liczby aktywnych użytkowników oraz ich efektywność, a także pośrednio przez zbadanie przyrostu danych OSM dotyczących obiektów naturalnych, historycznych i kolejowych.

Analiza aktywnych użytkowników OSM wykazała, największą ich liczbę w dwóch regionach Niemiec, co potwierdzają wcześniejsze badania opisane w literaturze (Neis, Zipf, 2012). Należy jednak zwrócić uwagę na znaczny przyrost liczby aktywnych użytkowników w Belgii, a także stosunkowo dużą ich liczbę na 1 mln ludności w Chorwacji. Wykonane badania potwierdziło heterogeniczność zarówno liczby użytkowników, jak i ich średniego przyrostu w czasie, w badanych regionach Europy. Przyjęte półroczne okresy są pewnego rodzaju generalizacją danych o liczbie aktywnych użytkowników w czasie, co może być uszczegółowione w przyszłych badaniach przez ich skrócenie do kwartałów lub miesięcy. Ostatecznie stwierdzono, że trend tempa przyrostu nowych aktywnych użytkowników OSM jest dodatni dla wszystkich badanych regionów i świadczy o zwiększaniu świadomości społeczeństwa dotyczącej danych przestrzennych, co może być dobrym początkiem tworzenia społeczeństwa geoinformacyjnego.

Przeprowadzona analiza efektywności użytkowników OSM pozwala jedynie na wyciągnięcie ogólnych i przybliżonych wniosków. Wynika to z faktu, iż z danych nie usunięto tych, które zostały zaimportowane z zewnętrznych baz, co powinny uwzględniać przyszłe badania. Ponadto analiza oparta jedynie na obiektach o wybranych kluczach powinna być poprzedzona oceną jakości danych. W przeprowadzonych badaniach wykryto sytuację błędnego przypisania tagu do obiektów przez jednego z użytkowników, co przełożyło się na błędne określenie efektywności dla całego analizowanego regionu. Przyjęta miara efektywności, którą stanowiła średnia liczba obiektów/długość/powierzchnia stworzona przez jednego aktywnego użytkownika danego rodzaju obiektów pozwala stwierdzić, że ogólnie bardziej efektywni są użytkownicy z regionów, w których ich sumaryczna liczba jest mniejsza. Nie można natomiast wskazać obszaru o najbardziej efektywnych użytkownikach, gdyż jak wykazały przeprowadzone badania jest ona zależna zarówno od reprezentacji geometrycznej, jak i kategorii (wyrażonej przypisanym kluczem) obiektów. To z kolei wskazuje na ściśle skorelowanie efektywności mierzonej w zaproponowany sposób z charakterem badanego obszaru, dlatego nie powinny być do niej wykorzystywane obiekty jednoznacznie wskazujące na ten charakter, przykładowo obiekty naturalne lub budynki. W przyszłości cennych wniosków może dostarczyć analiza rozmiaru danych wyrażonego w kilobajtach (kB) tworzonych przez aktywnego użytkownika oraz analiza efektywności w czasie. Na przykład w Brandenburgii, dla poligonowych obiektów naturalnych: w pierwszej połowie 2013 roku 169 użytkowników stworzyło obiekty o powierzchni 18 798 ha, natomiast rok później 200 użytkowników stworzyło obiekty o powierzchni 138 162 ha, podczas gdy w kolejnym półroczu było to zaledwie 73 497 ha przy tej samej liczbie użytkowników. Wskazuje to również na potrzebę głębszych badań nad motywacją użytkowników OSM.

Na podstawie analizy przyrostu danych OSM w czasie, można stwierdzić, że w okresie 01.2013-06.2016 r. najaktywniejsi byli użytkownicy w Słowacji i Brandenburgii. Dla obu tych regionów trend tempa wzrostu danych jest dodatni dla pięciu z ośmiu analizowanych zestawów danych. Podczas gdy dla regionu Nord-Pas-de-Calais tylko dla jednego zestawu wystąpił dodatni trend tempa wzrostu. W przyszłych badaniach analiza tempa wzrostu danych OSM powinna być przeprowadzona na całym zbiorze danych OSM, gdyż z dużym prawdopodobieństwem przyrost liczby obiektów o konkretnym tagu jest ściśle skorelowany z charakterem analizowanego regionu. Przyjęta metodyka miała za zadanie wyłonić temat danych, w którego tworzenie użytkownicy są bardziej zaangażowani, jednak otrzymane wyniki nie wskazują go jednoznacznie. W większości analizowanych zestawów danych trend tempa wzrostu jest ujemny, co może być związane z brakiem obiektów, które można wprowadzić do bazy OSM (co jest mało prawdopodobne) lub brakiem zainteresowania użytkowników tworzeniem danych o wybranych w niniejszym badaniu tagach. W wytłumaczeniu tego zjawiska należy wykluczyć spadek zainteresowania tworzeniem społecznościowych danych przestrzennych, gdyż liczba aktywnych użytkowników OSM we wszystkich badanych regionach wzrasta.

Literatura

- Andersen P., 2007: What is Web 2.0?: ideas, technologies and implications for education, vol. 1, no. 1: 1-64, JISC, Bristol, UK.
- Arsanjani J.J., Vaz E., 2015: An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. *International Journal of Applied Earth Observation and Geoinformation* 35: 329-337.

- Arsanjani J.J., Vaz E., Bakillah M., Mooney P., 2014: Towards initiating OpenLandMap founded on citizens' science: The current status of land use features of OpenStreetMap in Europe. Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Hiszpania.
- Arsanjani J.J., Zipf A., Mooney P., Helbich M., 2015: An introduction to OpenStreetMap in Geographic Information Science: Experiences, research, and applications. *OpenStreetMap in GIScience* :1-15, Springer International Publishing.
- 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* t. 10, z. 7(57): 15-24, PTIP, Warszawa.
- Cichociński P., Dębińska E., 2012: Badanie dostępności komunikacyjnej wybranej lokalizacji z wykorzystaniem funkcji analiz sieciowych. *Roczniki Geomatyki* t. 10, z. 4(54): 41-48, PTIP, Warszawa.
- Da Costa J.N., Bielecka E., Calka B., 2016: Jakość danych OpenStreetMap – analiza informacji o budynkach na terenie Siedleckizny, *Roczniki Geomatyki* t. 14, z. 2 (72): 201-211, PTIP, Warszawa.
- Girres J.F., Touya G., 2010: Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* 14: 435-459.
- Goodchild M.F., 2007: Citizens as sensors: the Word of volunteered geography. *GeoJournal* vol. 69.
- 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* vol. 37: 682-703.
- Hudson-Smith A., Batty M., Crooks A., Milton R., 2008: Mapping for the masses: Accessing web 2.0 through crowdsourcing. *Soc. Sci. Comput. Rev.* 27: 524-538.
- Lin Y. W., 2011: A qualitative enquiry into OpenStreetMap making. *New Review of Hypermedia and Multimedia* 17(1): 53-71.
- 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* t. 13, z. 3(69): 239-253, PTIP, Warszawa.
- Neis P., Zielstra D., Zipf A., 2012: The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007-2011. *Future Internet* 4: 1-21.
- 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 Int. J. Geo-Inf.* 1: 146-165.
- Schilling A., Over M., Neubauer S., Neis P., Walenciak G., Zipf A., 2009: Interoperable Location Based Services for 3D Cities on the Web Using User Generated Content from OpenStreetMap. Proceedings of the 27th Urban Data Management Symposium 2009, Słowenia.
- 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): 13-18, Leuven.
- Schmidt M., Klettner S., Steinmann R., 2013: Barriers for contributing to VGI projects. Proc. ICC, vol. 13.
- Steinmann R., Häusler E., Klettner S., Schmidt M., Lin Y., 2013: Gender Dimensions in UGC and VGI: A Desk-Based Study. Jekel/Car/Griesebner (Eds.): GI_Forum 2013 Creating the GISociety, Niemcy.
- Vickery G., Wunsch-Vincent S., 2007: Participative web and user-created content: Web 2.0 wikis and social networking. Organization for Economic Cooperation and Development (OECD).
- Yang A., Fan H., Jing N., Sun Y., Zipf A., 2016: Temporal analysis on contribution inequality in OpenStreetMap: A comparative study for four countries. *ISPRS International Journal of Geo-Information* 5(1), 5.
- 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 podjęto próbę oceny zaangażowania europejskiego społeczeństwa w tworzenie danych przestrzennych na przykładzie projektu OpenStreetMap (OSM). Jest to najbardziej popularny projekt działający w oparciu o crowdsourcing (ang. crowd – tłum, ang. sourcing – czerpanie), liczący niemal 3 miliony użytkowników na całym świecie i umożliwiający im tworzenie danych przestrzennych. OSM jest również największym źródłem danych dla naukowców zajmujących się tematyką volunteered geographic information (VGI), co potwierdzają badania literaturowe. Paneuropejską analizę zaangażowania społeczeństwa w tworzenie danych przestrzennych zbadano dla dziesięciu regionów odpowiadających poziomowi NUTS 1 Klasyfikacji Jednostek Terytorialnych do Celów Statystycznych, dla okresu od stycznia 2013 do czerwca 2016 roku w półrocznych interwałach. Wybór regionów poprzedzono analizą podobieństwa cech potencjalnie mających wpływ na tworzenie geodanych. Ostatecznie wybrane obszary to Chorwacja, Słowacja, Łotwa, Belgia oraz Region Centralnych i Wschodni (Polska), Brandenburgia i Dolna Saksonia (Niemcy) oraz Nord-Pas-de-Calais i Île de France (Francja). Porównania zaangażowania ludności w tworzenie społecznościowych danych przestrzennych dokonano w odniesieniu do trzech aspektów – wzrostu liczby aktywnych użytkowników OSM, oceny efektywności użytkowników i wzrostu ilości danych OSM w czasie. Przeprowadzone badania wskazują na wzrost zaangażowania użytkowników w tworzenie danych przestrzennych w wybranych regionach przy czym największy jest on w Brandenburgii i Dolnej Saksonii.

Abstract

The article attempts to assess the involvement of the European society in the development of spatial data on the example of the OpenStreetMap (OSM) project. It is the most popular Internet service operating on the basis of crowdsourcing, numbering almost 3 million users around the world and allowing them to create spatial data. OSM is also the largest source of data for scientists dealing with issues of volunteered geographic information (VGI), which is confirmed by research literature. The Pan-European analysis of public involvement in the development of spatial data was performed for ten regions corresponding to NUTS level 1 Nomenclature of Territorial Units for Statistics for the period from January 2013 to June 2016, in six-month intervals. The choice of regions was preceded by an analysis of their similarity, having the potential impact on geo-data creation. Finally, the selected areas are Croatia, Slovakia, Latvia, Belgium, the Central and Eastern Region (Poland), Brandenburg, the Lower Saxony (Germany) and the Nord-Pas-de-Calais and Île de France (France). The public involvement in the development of social spatial data has been compared with regard to three aspects – the increased number of active OSM users, evaluation of the effectiveness of users and the increased volume of OSM data in time. The studies show the increased involvement of users in the development of spatial data in selected regions with the highest results in in Brandenburg and the Lower Saxony.

mgr inż. Sylwia Marczak
sylwia.marczak1@gmail.com