

## OCENA PRZYDATNOŚCI OPENSTREETMAP JAKO ŹRÓDŁA DANYCH DLA ANALIZ SIECIOWYCH\*

### ASSESSMENT OF OPENSTREETMAP SUITABILITY AS A DATA SOURCE FOR NETWORK ANALYSIS

**Piotr Cichociński**

Katedra Geomatyki, Wydział Geodezji Górniczej i Inżynierii Środowiska  
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

**Słowa kluczowe:** jakość danych, normalizacja, dane przestrzenne, OpenStreetMap, analizy sieciowe

Keywords: data quality, standardization, spatial data, OpenStreetMap, network analysis

## Wstęp

Sieć jest to system połączonych ze sobą obiektów liniowych, poprzez które przesyłane są zasoby lub w oparciu o które odbywa się ruch (Husdal, 1999). Model danych sieci jest abstrakcyjną reprezentacją elementów składowych i cech rzeczywistych systemów sieciowych. Jednym z zastosowań analiz sieciowych w zakresie planowania transportu jest znajdowanie tras najkrótszych lub o najmniejszym koszcie pomiędzy dwoma lub większą liczbą punktów lub poszukiwanie miejsc położonych w określonym zakresie sumarycznego kosztu od wybranego punktu.

Model sieci może być zapisany w postaci grafu, który składa się z krawędzi reprezentujących liniowe kanały przepływu i węzłów reprezentujących miejsca połączeń pomiędzy krawędziami. W przypadku sieci drogowej krawędzie reprezentują osie dróg i ulic, a węzły odpowiadają skrzyżowaniom. Aby sieć działała jak model świata rzeczywistego, to z krawędziami musi być powiązana dodatkowa informacja określająca dopuszczalne kierunki ruchu oraz pewna wartość określająca koszt (albo opór) ruchu wzdłuż poszczególnych odcinków sieci.

Przeprowadzenie analiz sieciowych wymaga zgromadzenia i odpowiedniego przygotowania danych. W związku dużą popularnością odbiorników GPS i szerokim ich zastosowaniem w nawigacji samochodowej istnieje wiele firm oferujących zbiory danych sieciowych.

---

\* Praca zrealizowana w ramach badań statutowych prowadzonych w roku 2012 w Katedrze Geomatyki Wydziału Geodezji Górniczej i Inżynierii Środowiska Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie.

Jednak nabycie takich danych wiąże się ze sporymi kosztami. Dlatego interesującą alternatywą może być tutaj OpenStreetMap (Haklay, Weber, 2008) – projekt społecznościowy mający na celu utworzenie edytowalnej i dostępnej bez ograniczeń mapy świata. Mapa taka tworzona jest na podstawie danych z ręcznych odbiorników GPS, zdjęć lotniczych oraz innych dostępnych źródeł danych, a także szkiców wykonywanych w terenie. Ze względu na sposób budowania takiego zbioru, dane pochodzące z tego źródła za każdym razem wymagają jednak weryfikacji, poprawienia wychwyconych błędów i uzupełnienia braków.

## Charakterystyka OpenStreetMap

Projekt został utworzony, ponieważ większość map, które powszechnie uważane są za bezpłatne, w rzeczywistości posiada prawne lub techniczne ograniczenia co do ich użycia. Warunkiem stawianym danym dodawanym do bazy OpenStreetMap (OSM) jest aby były one poprawne, weryfikowalne i nie były objęte prawem autorskim lub też żeby osoba wprowadzająca dane miała do nich pełne prawa.

Dane zapisywane są w bazie OSM w postaci etykietowanych prostych elementów geometrycznych (*geometric primitive*). Należą one do jednego z trzech typów: punktów (*node*), linii (*way*) i relacji (*relation*). Obiekty powierzchniowe są reprezentowane przez zamknięte linie (pierwszy punkt jest identyczny z ostatnim). Bardziej skomplikowane struktury (na przykład obszary z enklawami) konstruowane są przy pomocy relacji. Z każdym z tych typów mogą być związane etykiety (*tag*) przyjmujące postać par klucz-wartość (*key-value pair*) i pełniące funkcję atrybutów. Istnieje dokładna lista akceptowanych etykiet i ich wartości, która może być rozbudowywana na drodze głosowania członków społeczności, lecz chociaż nie jest to zalecane, dopuszczalne jest używanie etykiet i wartości spoza tej listy.

Do września 2012 r. zasoby OpenStreetMap udostępniane były na podstawie licencji Creative Commons Attribution-ShareAlike 2.0 (Uznanie autorstwa-Na tych samych warunkach – CC-BY-SA), która stwierdza, że wolno kopiować, rozpowszechniać, odtwarzać i wykonywać utwór oraz tworzyć utwory zależne pod warunkiem oznaczenia w sposób określony przez twórcę lub licencjodawcę oraz rozpowszechniać nowopowstały utwór tylko na podstawie takiej samej licencji. Natomiast obecnie obowiązuje Open Database License (ODbL). Jej postanowienia są podobne do wcześniejszych, wystarczy jedynie zamienić „utwór” na „bazę danych”. Zasadniczą różnicą jest rozdzielenie danych użytych do wykonania mapy od formy ich prezentacji, ponadto zmodyfikowane dane OSM muszą być udostępniane wraz z mapą. Licencji na wykorzystanie danych udziela Fundacja OSM, a nie jak to było wcześniej poszczególne osoby, które wprowadziły dane do bazy.

Ponieważ baza ta budowana jest przez wolontariuszy, nie są sformułowane żadne plany jej systematycznego rozwoju. Dodanie nowych danych uzależnione jest od chęci poszczególnych osób do wykonania odpowiedniego pomiaru w terenie lub też wektoryzacji dostępnych zdjęć lotniczych. Zdarzają się jednak przypadki, że firmy bądź instytucje posiadające w swoich zasobach różne dane decydują się przekazać je społeczności OpenStreetMap nieodpłatnie. W skali świata znaczące są zezwolenia jakie udzieliły firmy Yahoo i Microsoft na nieograniczone wykorzystanie danych (w szczególności zdjęć lotniczych) prezentowanych na ich portalach mapowych. W Polsce już kilka samorządów zdecydowało się udostępnić na potrzeby OSM posiadane w swoich zasobach dane: Siedlce (Zaborowski, 2010), Szczecin (Zaborowski, 2011), Wrocław, Bytom, Police i Łódź (Czernik, 2012).

## Jakość w zakresie informacji geograficznej

Poświęcona podstawom systemów zarządzania jakością norma ISO 9000 (PN-EN ISO 9000, 2006) definiuje jakość jako całość charakterystyk produktu, która zależy od możliwości zaspokojenia przez niego wyrażonych i ukrytych potrzeb. W skrócie można stwierdzić, że jakość informuje o przydatności do wykorzystania.

W zakresie geomatyki opis jakości ma dwa podstawowe cele:

- umożliwienie twórcom informacji geograficznej definiowania jak dobrze ich produkt spełnia specyfikacje,
- pomoc użytkownikom w określeniu wymagań wobec danych.

Jakość należy rozpatrywać pod czterema względami (PN-EN ISO 19113, 2002):

- historycznym – pochodzenie zbioru danych geograficznych, obejmujące opis sposobu utworzenia i historii zbioru danych geograficznych: celu produkcji, czasu powstania, wykorzystanych materiałów źródłowych, zastosowanych przetworzeń oraz odpowiedzialnych za nie organizacji,
- wcześniejszego wykorzystania – kto, kiedy i w jakim celu użytkował określony zbiór danych; najistotniejszą w tym kontekście informacją jest jakie wykrył błędy i jakie ograniczenia zostały narzucone lub zauważone w trakcie użytkowania,
- możliwej niejednorodności miar jakościowych w tym samym zbiorze danych geograficznych spowodowanej zróżnicowanym rozkładem uzależnionym od zasięgu przestrzennego, zasięgu czasowego, typów encji, wartości atrybutów itp.,
- ilościowego pomiaru względem wzorca jakości – wzorzec taki nazywany jest przestrzenią rozważań (ang. *nominal ground, universe of discourse*) – jest to model reprezentujący określone spojrzenie na świat rzeczywisty, idealny zbiór danych geograficznych.

Ocenę jakości przeprowadza się jedną z dwóch metod: pośrednią lub bezpośrednią (PN-EN ISO 19114, 2005). Pośrednia metoda oceny oparta jest na wiedzy zewnętrznej, takiej jak pochodzenie i wykorzystanie zbioru, jakość danych użytych do wytworzenia zbioru. Zalecana jest tylko i wyłącznie wtedy, gdy nie można wykorzystać metod oceny bezpośredniej. Ocena bezpośrednia może polegać na testowaniu każdej jednostki w populacji określonej przez zakres jakości danych (pełna kontrola) lub też dotyczyć wybranych jednostek w liczbie wystarczającej do otrzymania wyniku (próbkiwanie). W tym drugim przypadku istotne jest właściwe określanie rozmiaru próbki, co w przypadku danych geograficznych jest trudniejsze niż dla danych tabelarycznych. Rozmiar próbki może być związany z liczbą obiektów danego typu, obszarem obejmowanym przez zbiór danych, całkowitą długością krzywych w zbiorze danych, całkowitą liczbą wierzchołków opisujących krzywe lub obszary w zbiorze danych.

Szczególny problem związany jest z dynamicznymi (czyli stale zmieniającymi się) zbiorami danych. Można go rozwiązać na dwa sposoby: albo w regularnych odstępach czasu dokonywać kontroli wykonanej kopii jak zbioru statycznego lub też dokonywać ciągłej kontroli – oceny jakości danych aktualizujących i ich wpływu na zbiór danych.

Metoda oceny bezpośredniej ma dwa warianty:

- wewnętrzny – gdy wszystkie niezbędne dane do takiej oceny znajdują się w ocenianym zbiorze danych (przykładem może być ocena spójności topologicznej),
- zewnętrzny – gdy wymagane są dodatkowe (zewnętrzne) dane odniesienia (przykładami mogą być: ocena kompletności nazw ulic, ocena dokładności położenia).

O zgodności z wzorcem (przestrzenią rozważań) informuje pięć następujących elementów jakości:

1. Dokładność położenia – parametr ten opisuje dokładność określenia współrzędnych obiektu. Wyrażany jest z reguły przy pomocy średniego błędu kwadratowego lub elipsy błędu średniego. W praktyce preferowaną metodą określenia dokładności położenia jest porównanie z niezależnym źródłem o większej dokładności.
2. Dokładność tematyczna – czynnik ten opisuje dokładność lub pewność pozyskania wartości atrybutu. Oszacowanie dokładności atrybutów ilościowych jest analogiczne do określania dokładności położenia. Jego wynikiem jest odchylenie standardowe wartości atrybutu. W przypadku atrybutów o charakterze jakościowym określa się pewność wartości atrybutu.
3. Aktualność – opisuje moment lub okres czasu, w którym zawartość bazy danych odpowiada rzeczywistości. Można ją reprezentować przy pomocy następujących parametrów: daty powstania obiektu, daty pomiaru, daty wprowadzenia do bazy, średniego okresu czasu pomiędzy pomiarem a wprowadzeniem do bazy, szybkości zmiany obiektów lub ich atrybutów.
4. Kompletność – raport dotyczący kompletności powinien opisywać związki między przedstawionymi w bazie danych obiektami a modelowaną przy ich pomocy rzeczywistością. W szczególności raport powinien opisywać na ile wyczerpujący jest zbiór obiektów.
5. Spójność logiczna – raport na temat spójności logicznej powinien opisywać wierność związków zapisanych w strukturze cyfrowych danych przestrzennych.

## **Jakość OpenStreetMap w innych krajach europejskich**

Najbardziej kompleksowa ocena dokładności OpenStreetMap została przedstawiona w pracy (Girres, Touya, 2010) i dotyczyła obszaru Francji. Został przeanalizowany szeroki zakres elementów jakościowych, takich jak dokładność geometryczna, atrybutowa, semantyczna i czasowa, spójność logiczna, kompletność, pochodzenie i wykorzystanie. Wiele z wykonanych działań polegało na porównaniu z instytucjonalnymi danymi pochodzącymi z francuskiego Institut Géographique National (IGN).

Jak stwierdzili autorzy powyższego opracowania, było ono rozszerzeniem prac zaprezentowanych wcześniej w publikacji (Haklay, 2010) poświęconej porównaniu dokładności OSM z mapami brytyjskiej Ordnance Survey (OS). Haklay założył, biorąc pod uwagę obowiązujące procedury zapewnienia jakości, że zbiory danych Ordnance Survey charakteryzują się wyższą dokładnością (co najmniej w zakresie położenia i atrybutów). Nie bez znaczenia był także ustawowy wymóg, żeby 99,6% istotnych obiektów świata rzeczywistego zostało zaprezentowanych w bazie danych w przeciągu 6 miesięcy od ich powstania.

Inne podejście do zagadnienia oceny dokładności zostało zaprezentowane w pracy (Mondzech, Sester, 2011). Autorki przeanalizowały dane OpenStreetMap pod względem możliwości zastosowań, skupiając się na wyznaczaniu optymalnych tras dla ruchu pieszego. W tym celu dokonały porównania z trasami wyznaczonymi na podstawie funkcjonującej w Niemczech bazy danych topograficznych ATKIS. Dostrzegły przy tym zasygnalizowany również w pracy (Cichociński, Dębińska, 2012) problem placów rejestrowanych w bazie OSM jako obiekty powierzchniowe i zaproponowały jego rozwiązanie.

Jednak, jak pokazują kolejne badania, nie zawsze inne zbiory, względem których chciałoby się dokonać porównania można uznać za bardziej dokładne. Pokazują to wyniki badań zaprezentowane w pracach (Ciepluch i in., 2010) i (Zielstra, Zipf, 2010). Pierwsza z tych publikacji pokazuje wyniki ręcznego porównania dokładności OSM z popularnymi produktami komercyjnymi: Google Maps i Microsoft Bing Maps dla obszaru Irlandii. Do analizy wybrano stolicę, małe i duże miasto oraz miasteczko uniwersyteckie. Zbadano kompletność, aktualność oraz poprawność względem rzeczywistości, opierając się na terenowej weryfikacji danych. Okazało się, że żadna z trzech przebadanych baz danych nie charakteryzowała się spójną jakością i w każdej z nich można było znaleźć ewidentne błędy. Natomiast Zielstra i Zipf stwierdzają, że przynajmniej na obszarach dużych niemieckich miast OpenStreetMap może być interesującą alternatywą dla, jak to określają, „profesjonalnych” zbiorów danych.

## **Analiza wybranych elementów jakości danych OpenStreetMap dla obszaru Polski**

### **Kompletność – nadmiar i niedomiar**

Podobnie jak dla większości elementów jakościowych, zbadanie zbioru danych pod względem kompletności wymaga posiadania danych referencyjnych lub dokonania porównania z rzeczywistością. W przypadku sieci drogowej dodatkowym utrudnieniem może być niezgodność klasyfikacji obiektów. Chociaż zalecane jest opisywanie dróg krajowych jako *primary*, a wojewódzkich jako *secondary*, to jednak na tym poziomie ważności dróg nie ma problemu z kompletnością, gdyż twórcy dostrzegają istotność tej grupy danych dla zastosowań w zakresie nawigacji samochodowej. Problem natomiast pojawia się przy drogach mniej ważnych, które nie mają swoich odpowiedników w oficjalnych bazach danych. Dlatego jednoznaczne stwierdzenie na tym poziomie nadmiaru bądź niedomiaru jest praktycznie niemożliwe.

Autor pokusił się jedynie o sprawdzenie ubytków wywołanych zmianą licencji. Twórcy, którzy wprowadzali swoje dane w okresie obowiązywania licencji CC-BY-SA, byli proszeni o zaakceptowanie nowej licencji ODbL. W przypadku, gdy zgoda nie została wyrażona, odpowiednie dane zostały usunięte. Polska, jako jeden z kilku krajów, szczególnie ucierpiała na tej operacji. W literaturze można znaleźć metodę oceny kompletności obiektów liniowych polegającą na porównaniu sumarycznej długości wszystkich odcinków w całej bazie danych lub w poszczególnych oczkach sztucznie wygenerowanej siatki (Ciepluch i in., 2011).

Autor natomiast proponuje do rozwiązania tego zadania zastosowanie jądrowego estymatora gęstości, który wyznacza gęstość występowania obiektów punktowych lub liniowych w pobliżu każdej komórki wynikowego rastra. Realizowane jest to w ten sposób, że nad każdym obiektem jest wpasowywana łagodnie zakrzywiona powierzchnia. Jej wartość jest najwyższa dokładnie nad obiektem i maleje wraz ze wzrostem odległości od niego, osiągając wartość zero w odległości określonej przez parametr zasięg poszukiwań. Objętość przestrzeni pod powierzchnią wynosi 1. Gęstość dla każdej komórki rastra wynikowego wyliczana jest przez dodanie wartości poszczególnych powierzchni znajdujących się nad środkiem tej komórki. W programie ArcGIS, którego użyto do przygotowania mapy gęstości

zaprezentowanej na rysunku 1, stosowana jest funkcja kwadratowa, zwana inaczej funkcją Epanecznikowa. Krytycznym parametrem wpływającym na ostateczny wynik jest promień przeszukiwania. Jego mniejsza wartość powoduje zaakcentowanie, ale również ograniczenie pod względem zajmowanej powierzchni lokalizacji o podwyższonej gęstości, natomiast wartość większa daje w wyniku model bardziej wygładzony. Ostatecznie zdecydowano się na promień przeszukiwania o wielkości 10 000 m. Ze strony <http://download.geofabrik.de/> pobrano w postaci plików *shape* (Esri, 1998) dwie wersje danych sieci drogowej dla Polski: jedną z okresu tuż przed usunięciem danych z niezgodną licencją i drugą obrazującą stan bazy danych dwa miesiące po dokonaniu drastycznych zmian. Obliczono gęstość dla każdej z wersji, a następnie odjęto od siebie wynikowe rastry. Rysunek 1 prezentuje ostateczny wynik przeprowadzonej analizy. Uzyskany rezultat pokazuje, że są miejsca gdzie danych rzeczywiście znacząco ubyło, ale też można znaleźć obszary, na których widać pozytywny wpływ zagrożenia obniżeniem jakości bazy danych na aktywność twórców.

### Spójność logiczna – topologiczna

Istotnym ograniczeniem w wykorzystaniu różnych zbiorów danych dla celów analiz sieciowych są wzajemne relacje przestrzenne pomiędzy liniami – kiedy koniec jednej linii nie styka się precyzyjnie z końcami innych linii w danym zbiorze. Ponadto, jeżeli linie przecinają się ze sobą, należy określić czy jest to zamierzone, czy też powinno być potraktowane jako błąd. Często oprogramowanie GIS „naprawia” wszelkie przecięcia linii, dokonując ich podziału na mniejsze części, stykające się ze sobą końcami. Powstaje w ten sposób graf płaski. Jednak są sytuacje, gdy jest to niepożądane, na przykład przy modelowaniu skrzyżowań wielopoziomowych (Fischer, 2003). Dlatego każdy taki przypadek powinien być indywidualnie rozpatrzony. Przypadki drobnych niedokładności w powiązaniu ze sobą linii, wynikające na przykład z zaokrąglenia współrzędnych mogą być automatycznie poprawione przy założeniu, że jeżeli dwa końce linii są od siebie odległe nie więcej niż o pewną wartość tolerancji, to powinny być uznane za tożsame i z pewnością tworzą jeden węzeł.

O ile inne elementy jakościowe wymagają w praktyce pracochłonnego porównania z rzeczywistym przebiegiem dróg i ulic, o tyle spójność geometryczną można zbadać metodami oceny wewnętrznej. W rozwiązaniu tego problemu może pomóc topologia wykorzystana do opisu przestrzennych związków pomiędzy obiektami geograficznymi (Cichociński, 2007).

W przypadku takich zbiorów obiektów jak sieć drogowa, w celu sprawdzenia spójności geometrycznej wystarczy skorzystać z dwóch reguł topologicznych: „linie nie mogą się przecinać i nakładać na siebie”, „linie muszą się stykać ze sobą wyłącznie końcami”. Po zasygnalizowaniu wykrytych w ten sposób błędów, można je poprawić zarówno metodami automatycznymi (w przypadku przecięć), jak też metodami półautomatycznymi bądź ręcznymi (w przypadku błędów węzłów wiszących).

Innym praktycznym rozwiązaniem jest skorzystanie z programu *spatialite\_osm\_net*, który jest jednym z narzędzi towarzyszących systemowi zarządzania bazą danych przestrzennych *Spatialite* (Krawczyk, 2012). Przetwarza on plik XML z danymi *OpenStreetMap* i zapisuje w tablicy bazy danych przestrzennych wybrane obiekty z etykietą *highway*, będące elementami sieci drogowej, po których może się odbywać ruch pojazdów. Ponadto dodawany jest atrybut opisujący koszt ruchu po poszczególnych odcinkach sieci, wyliczany na podstawie arbitralnie przyjętych maksymalnych prędkości dla poszczególnych typów dróg (*motorway*, *primary*, *secondary*, etc.). Najważniejszym jednak efektem pracy programu jest

poprawienie znaczącej większości błędów topologicznych występujących w danych, z zachowaniem jednak takich krytycznych miejsc jak skrzyżowania wielopoziomowe. Istotna jest także zamiana obiektów powierzchniowych takich jak ronda, na linie tworzące ich obwód, bo tylko w takiej postaci mogą być wykorzystane w analizach sieciowych.

### Dokładność położenia

W przypadku obiektów punktowych łatwo i jednoznacznie można określić odległość pomiędzy lokalizacją badaną i referencyjną. Dla obiektów liniowych zadanie nie jest już takie proste do wykonania. W literaturze można znaleźć różne propozycje określania różnic w położeniu takich obiektów. Do bardziej popularnych miar można zaliczyć odległość Hausdorffa (Hanguet, 1995) oraz część całkowitej długości obiektu o mniejszej dokładności, która znajduje się w określonej odległości od obiektu o wyższej dokładności (Goodchild, Hunter, 1997; Hunter, 1999).

Dostępne jest narzędzie, które posługując się opisanymi powyżej metodami pozwoli, chociaż nie jest to jego główne przeznaczenie, na ocenę dokładności położenia obiektów liniowych. Tym narzędziem jest oprogramowanie RoadMatcher (Vivid Solutions, 2005). Zostało opracowane przez kanadyjską firmę Vivid Solutions na zlecenie rządu prowincji Kolumbia Brytyjska, poszukującego wydajnego, wysoko zautomatyzowanego oprogramowania do przetwarzania danych przestrzennych. Dodatkową, oprócz funkcjonalności, zaletą jest jego przynależność do grupy wolnego oprogramowania, co umożliwia nie tylko bezpłatne użytkowanie, ale również modyfikację kodu źródłowego i dopasowanie go do swoich potrzeb (Michalak, 2007).

Głównym zadaniem tego programu jest znajdowanie odpowiadających sobie elementów w dwóch zbiorach danych sieciowych (rys. 2), a następnie łączenie ich celem utworzenia jednego spójnego wynikowego zbioru danych. Aby zapewnić dopasowanie jeden do jednego w niektórych przypadkach konieczny jest podział elementów na krótsze odcinki. Jeden z dopasowanych elementów, określony jako preferowany, staje się elementem odniesienia i zostaje włączony do wynikowej sieci. W przypadku braku dopasowania, dany element może zostać bezpośrednio przeniesiony do wyniku lub też odrzucony.

Program RoadMatcher wykorzystano do zbadania dokładności położenia dróg z obszaru podkrakowskiej gminy Zabierzów. Autor dysponuje informacjami na temat sieci drogowej z tego terenu, pozyskanymi poprzez wektoryzację osi dróg na mapach zasadniczych i ewidencyjnych. Dla tego samego obszaru zostały pobrane ze strony <http://openstreetmap.org> dane w formacie XML i przetworzone programem `spatialite_osm_net` do formatu *shapefile*.

Odpowiadające sobie elementy sieci są identyfikowane w programie RoadMatcher w określonym przez użytkownika promieniu poszukiwań. Rozpoczęto od domyślnej wartości 10 m. Ponieważ nie wszystkie pary zostały utworzone, a zaobserwowano odległości pomiędzy obiektami nawet kilkudziesięciometrowe, ostatecznie ustalono wielkość promienia poszukiwań na 60 m. Pierwotna liczba 1341 elementów w zbiorze OSM i 1891 w zbiorze referencyjnym zwiększyła się odpowiednio do 1772 i 2118, z powodu wymuszonego koniecznością lepszego dopasowania podziału na krótsze odcinki. W promieniu poszukiwań znalazło się w każdym zbiorze po 1188 obiektów o zbliżonym kształcie, a także 62 obiekty w zbiorze OSM i 67 w zbiorze referencyjnym, znacząco różniące się kształtem. Odpowiednio 522 i 863 obiekty zostały uznane za zupełnie różne i nie mające swoich odpowiedników w drugim zbiorze danych.

Dla 1188 par powiązanych ze sobą obiektów można było obliczyć odległość Hausdorffa. Minimalna wyznaczona odległość to 0,04 m, maksymalna 41,81 m, średnia 4,47 m, a odchylenie standardowe średniej 4,34 m. Tym samym potwierdzone zostało stwierdzenie Haklay'a (2010), który zauważył, że średnio odchyłka względem danych Ordnance Survey nie przekraczała 6 metrów.

Dlatego odległości 6 metrów użyto do wyznaczenia bliskości (*nearness*), czyli procentowo wyrażonego stopnia zawierania się całkowitej długości obiektu z bazy OSM w strefie buforowej o określonej wielkości wokół obiektu referencyjnego. Średnia wartość bliskości osiągnęła poziom 91%.

### Dokładność czasowa

Istotną cechą OpenStreetMap jest błyskawiczna reakcja na powstawanie nowych elementów sieci drogowej. W szczególności dotyczy to istotnych dróg, takich jak autostrady. Nowo otwarty odcinek jest prawie natychmiast widoczny na mapie. Jest to możliwe między innymi dzięki temu, że w bazie danych zapisywana jest informacja również o drogach w budowie i projektowanych. W związku z tym wystarczy tylko zmienić wartość etykiety *highway* z *construction* na właściwą.

## Wnioski

Zaprezentowane powyżej przykłady pokazują, że istnieją i dostępne są dla każdego zainteresowanego narzędzia umożliwiające kontrolę poprawności danych OpenStreetMap. Za ich pomocą relatywnie łatwe jest wykonanie oceny jakości metodami wewnętrznymi, nie wymagającymi posiadania innych danych. Metody oceny jakości z użyciem danych zewnętrznych też nie są nadmiernie skomplikowane, jednak możliwość ich stosowania jest ograniczona i uzależniona od dostępu do innych zbiorów o wyższej dokładności. Klóci się to również z ideą projektu OpenStreetMap, który powstał właśnie z powodu braku otwartego dostępu do instytucjonalnych zbiorów danych.

Za najcenniejsze można uznać wyniki przeprowadzonej oceny dokładności położenia. Pokazują one, że pomimo niewielkich średnich różnic występujących pomiędzy odpowiadającymi sobie reprezentacjami obiektów w dwóch porównywanych zbiorach danych, to jednak znacząca jest liczba błędów grubych, nie wynikających wcale z dostępnych dla twórców materiałów źródłowych.

### Literatura

- Cichociński P., 2007: Zastosowanie zaawansowanych reguł topologicznych w procesie budowania baz danych przestrzennych wspomagających wycenę nieruchomości. *Roczniki Geomatyki* t. 5, z. 3, PTIP, Warszawa.
- Cichociński P., Dębińska E., 2012: Badanie dostępności komunikacyjnej wybranej lokalizacji z wykorzystaniem funkcji analiz sieciowych. *Roczniki Geomatyki*, r. 10, z. 4, PTIP, Warszawa.
- Ciepluch B., Jacob R., Mooney P., Winstanley A., 2010: Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Leicester, UK, 20-23 July 2010.
- Ciepluch B., Mooney P., Winstanley A., 2011: Building Generic Quality Indicators for OpenStreetMap. Proceedings of the 19th annual conference of GIS Research UK (GISRUK 2011). Portsmouth, 27-29 April 2011.



- Czernik Z., 2012: Samorzady powoli zauważają OSM. [W:] Portal polskiej społeczności OpenStreetMap. <http://openstreetmap.org.pl/osm/2012/samorzady-powoli-zauwazaja-osm/>
- Esri, 1998: ESRI Shapefile Technical Description. An ESRI White Paper. Environmental Systems Research Institute. Redlands.
- Fischer M.M., 2003: GIS and Network Analysis. ERSA conference papers. 43rd European Congress of the European Regional Science Association. Jyväskylä, Finland, 27th-30th August 2003.
- Girres J.-F., Touya G., 2010: Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 2010, 14(4): 435-459.
- Goodchild M.F., Hunter G.J., 1997: A Simple Positional Accuracy Measure for Linear Features. *International Journal of Geographical Information Science*, 11(3): 299-306.
- Haklay M., Weber P., 2008: OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, October–December 2008: 12-18.
- Haklay M., 2010: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B*, 37(4): 682-703.
- Hanguet J.F., 1995: Computation of the Hausdorff distance between plane vector polylines. *Proceedings of AUTO-CARTO 12*, 4: 1-10.
- Hunter G.J., 1999: New tools for handling spatial data quality: moving from academic concepts to practical reality. *URISA Journal*, 11(2): 25-34.
- Husdal J., 1999: Network analysis – network versus vector – A comparison study. <http://www.husdal.com/1999/10/11/network-analysis-raster-versus-vector-a-comparison-study/>
- Krawczyk A., 2012: Możliwości wykorzystania personalnej bazy danych przestrzennych SpatialLite w praktyce górniczej. *Roczniki Geomatyki* t. 10, z. 2, 45-49, PTIP, Warszawa.
- Michalak J., 2007: Otwarte oprogramowanie i otwarte dane w geomatyce. *Roczniki Geomatyki* t.5, z. 2, PTIP, Warszawa.
- Mondzech J., Sester M., 2011: Quality Analysis of OpenStreetMap Data Based on Application Needs. *Cartographica* 46:2, 115-125.
- PN-EN ISO 19113, 2002: Informacja geograficzna – Podstawy opisu jakości
- PN-EN ISO 19114, 2005: Informacja geograficzna – Procedury oceny jakości
- PN-EN ISO 9000, 2006: System zarządzania jakością — Podstawy i terminologia
- Vivid Solutions, 2005: RoadMatcher. <http://www.vividsolutions.com/products.asp?catg=spaapp&code=roadmatcher>
- Zaborowski A., 2010: Szczecin najbardziej kompletnym miastem. [W:] OpenStreetMap i okolice. Wolne dane w praktyce. <http://blog.openstreetmap.pl/2010/szczecin-najbardziej-kompletnym-miastem>
- Zaborowski A., 2011: Hiszpański kataster uwolniony. [W:] OpenStreetMap i okolice. Wolne dane w praktyce. <http://blog.openstreetmap.pl/2011/hiszpański-kataster-uwolniony>
- Zielstra D., Zipf A., 2010: A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. *Proceedings of the 13th AGILE International Conference on Geographic Information Science*. 10-14 May 2010, Guimarães, Portugal.

### **Abstract**

*For several years now, GIS users have at their disposal the data set constituting some kind of alternative to commercial products. It is the OpenStreetMap (OSM) – a project with the objective to build an editable and available without restrictions map of the world. This map is created based on data from handheld GPS receivers, aerial photographs and other available data sources, as well as sketches made in the field. Acquired information is stored in one central database, from which it can be downloaded as a map image using selected symbols or as vector data.*

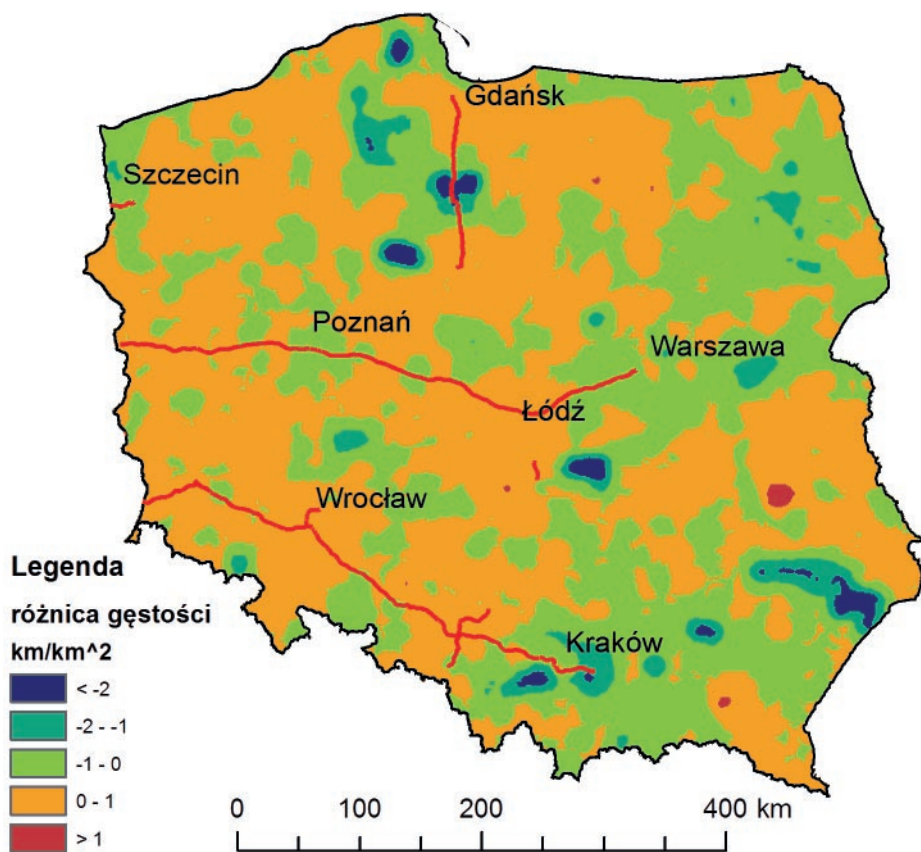
*Data from this source can be used, among others, for network analysis, of which the most widely used and the most common function is finding the optimal route between two points. For the result of this analysis to be considered reliable, the data used must be of appropriate quality.*

*Because the OSM database is built by volunteers, there are no plans of its systematic development and there is no top-down quality control. Therefore, the paper proposes methods and tools to verify the*

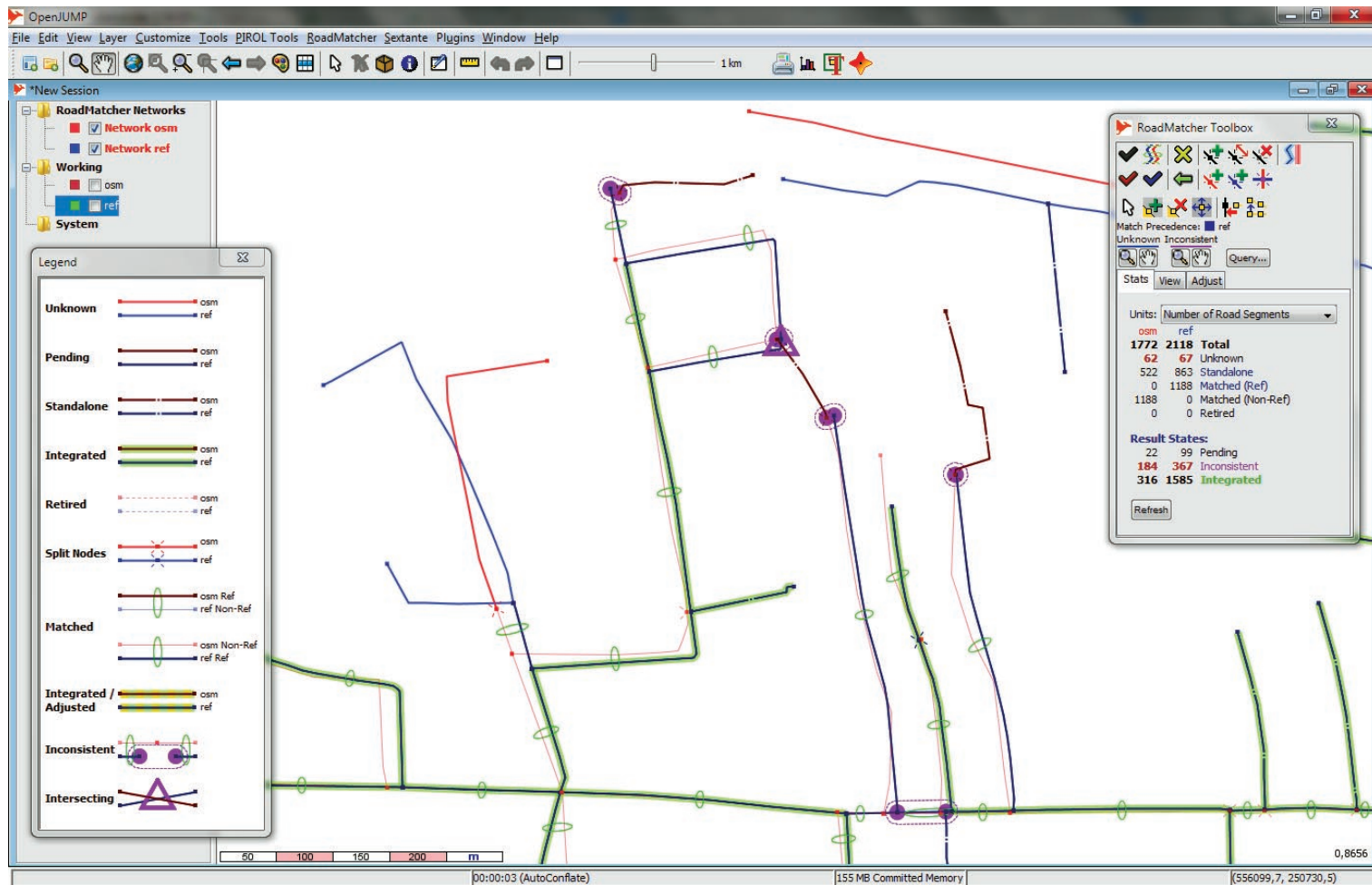
*suitability of the so far collected data for the purposes of network analysis, as well as to correct errors found. Among the quality aspects considered for geographical data, particular attention was paid to the completeness, positional accuracy, topological consistency and temporal accuracy.*

*The first two problems require in practice comparison with the actual course of roads and streets. To solve the problem of geometric consistency, the use of relevant topological rules examining selected spatial relationships between geographic objects was suggested. When the errors detected this way are marked, they can be corrected using automated methods, and also semi-automatic or manual methods.*

dr inż. Piotr Cichociński  
Piotr.Cichocinski@agh.edu.pl  
tel. (12) 617-34-31



**Rys. 1.** Różnica gęstości sieci drogowej w bazie OpenStreetMap dla obszaru Polski dla okresów sprzed i po zmianie licencji



Rys. 2. Wynik działania programu RoadMatcher