

Geodezja i kartografia wobec wyzwań Linked Open Data oraz internetu 3.0



LOD: na progu nowej ery

Zainicjowana kilka lat temu polityka otwartości danych mocno odmieniła branżę geodezyjną i kartograficzną. Prawdziwa rewolucja jednak dopiero przed nami. Liczy się bowiem nie tylko to, co uwalniamy, ale również w jaki sposób to robimy.

Jerzy Królikowski

Zróbmy krótki eksperyment. Wpiszmy w Google'u frazę „Ile lat ma Beta Kozidrak?”. Jeszcze kilka lat temu wyszukiwarka zapewne wyświetliłaby nam linki do portali plotkarskich, które zachwycają się, jak świetnie trzyma się ta wokalistka, a dziś po prostu dostajemy odpowiedź na zadane pytanie. A gdyby tak w podobny sposób można wykonywać analizy przestrzenne? Na przykład poprzez wpisanie frazy „Wskaż miasta powiatowe w Polsce powyżej 50 tys. mieszkańców leżące do 25 km od autostrady”. Na razie to dla Google'a zbyt wiele, ale szybki postęp, jaki dokonuje się wokół technologii rozwijanych pod hasłem Linked Open Data (LOD) czy Web 3.0, pozwala przypuszczać, że to wcale nieodległa przyszłość.

• Jak otwierać, to na całego

Nim przejdziemy do wyjaśniania obu wyżej podanych terminów, zatrzymajmy się na chwilę przy otwieraniu danych, bo to właśnie popularyzacja polityki Open Data była bodźcem do sformułowania idei LOD. Lawinę uwalniania rządowych zasobów uruchomili jeszcze w ubiegłym wieku Amerykanie. Rząd USA doszedł bowiem do wniosku, że skoro stworzono jakieś dane za pieniądze podatników, to zmuszanie ich do płacenia za nie po

raz drugi jest po prostu nieuczciwe. Idea ta powoli przebijała się do innych krajów, ale za moment przełomowy można uznać uwolnienie w 2010 roku zasobów brytyjskiej agencji kartograficznej Ordnance Survey. Przez lata twardo broniła ona bowiem poglądu, że jeśli dane mają być dobrej jakości, to trzeba za nie płacić.

Pięć kroków do otwartości

W 2006 roku Tim Berners-Lee – jeden z twórców internetu – opracował pięciostopniową skalę otwartości danych. Im wyższy poziom, tym uwolnione dane są bardziej dostępne:

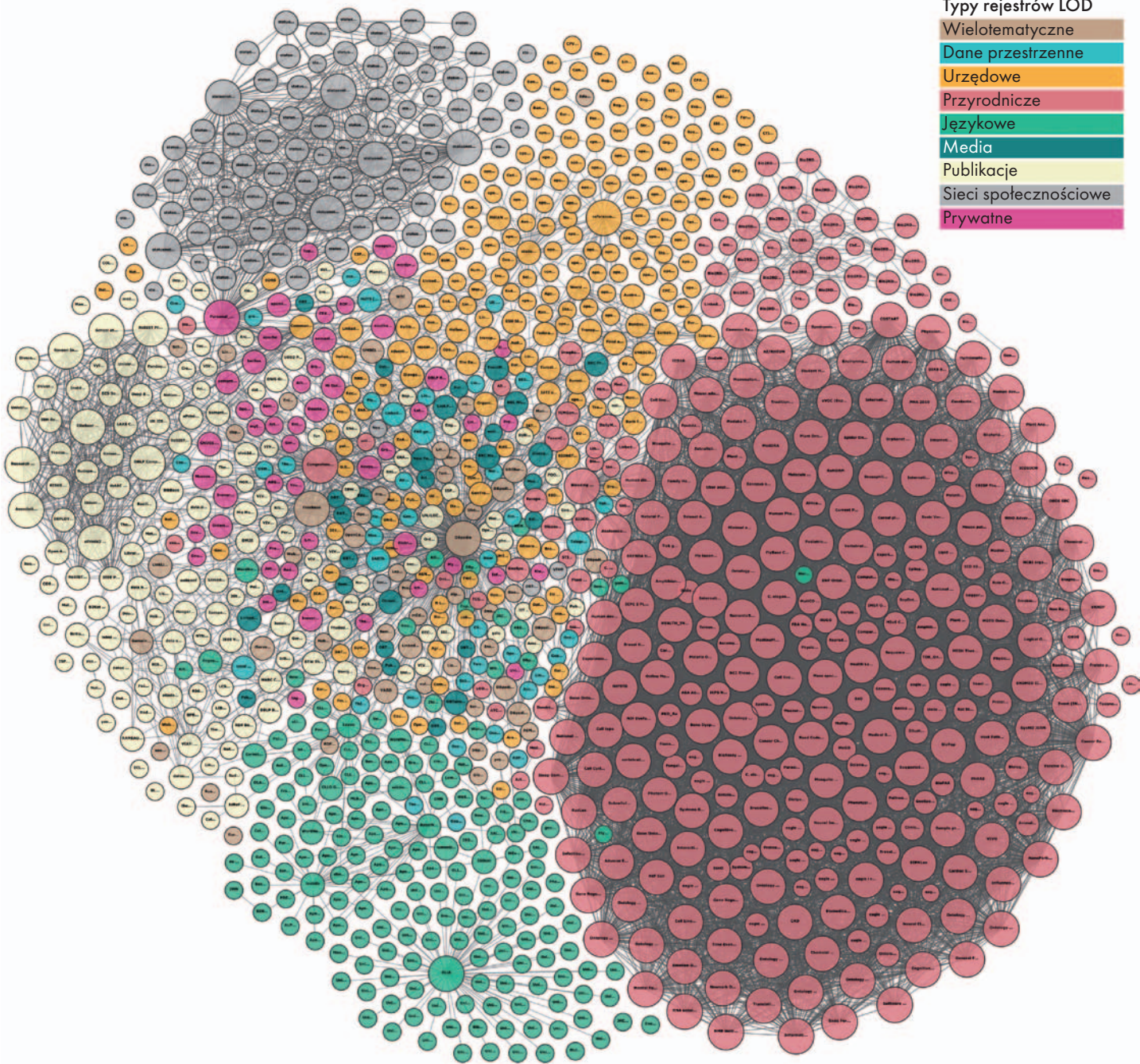
1. Dokument dostępny w internecie na wolnej licencji w formie nieustrukturyzowanych danych, np. jako skan dokumentu lub mapy.
2. Otwarte dane w formie ustrukturyzowanej, ale zapisanej w zamkniętej formie np. jako tabela w formacie XLS czy plik SHP.
3. Dane jak w punkcie 2, ale w niezabezpieczonym formacie, np. XLSX czy GML.
4. Dane zapisane w otwartym standardzie międzynarodowego konsorcjum W3C: RDF czy SPARQL, dzięki czemu inni użytkownicy mogą linkować do nich swoje zasoby.
5. Dane jak w pkt 4, ale połączone poprzez relacje.

Na naszym podwórku moda na uwalnianie geodanych rozpoczęła się od nowelizacji *Prawa geodezyjnego i kartograficznego* z 2014 roku otwierającej dostęp do czterech baz: obiektów ogólnogeograficznych, modeli terenu w niskiej rozdzielczości, nazw geograficznych oraz rejestru granic. Nie mniej spektakularne było uwolnienie w 2016 roku przez Państwowy Instytut Geologiczny skanów czterech serii map geologicznych – na tak daleko idący krok nie zdecydował się jeszcze żaden inny kraj.

Ten przyspieszający proces otwierania geodanych z błogosławieństwa powoli staje się przekleństwem. Po pierwsze, w gąszczu darmowych rejestrów coraz trudniej znaleźć interesujące nas materiały. Po drugie, nawet jeśli znajdziemy to, co nas interesuje, często niełatwo dobrać, jak dany plik pobrać, otworzyć i zinterpretować znajdujące się w nim dane. Trzecim problemem jest łączenie tych darmowych rejestrów i prowadzenie na ich podstawie analiz przestrzennych.

W tej sytuacji państwowe instytucje zaczęły dostrzegać, że samo wyłożenie swoich zasobów takich, jakimi są, to za mało, by obywatele, przedsiębiorcy i naukowcy chcieli z nich korzystać. Ważna jest również forma udostępniania. Weźmy przykład danych przestrzennych. To miło, że urząd opublikuje na swoich serwerach skan mapy, ale przy wykonywaniu analiz przestrzennych na niewiele

Wielotematyczne
Dane przestrzenne
Urzędowe
Przyrodnicze
Językowe
Media
Publikacje
Sieci społecznościowe
Prywatne



Fot. lod-cloud.net

Rys. 1. Graf prezentujący powiązania między istniejącymi bazami LOD

się on zda. Znacznie przydatniejsza będzie forma wektorowa, najlepiej zapisana w jakimś otwartym standardzie. Tak oto dochodzimy do trzeciego stopnia otwartości danych w skali Tima Bernersa-Lee (patrz ramka). Tu jednak się ona nie kończy. Berners-Lee uważa bowiem, że zbiór jest w pełni otwarty dopiero wtedy, gdy połączymy jego zawartość z innymi rejestrami. Przecież jeden obiekt (rzeka, osoba czy miejscowość) obecny jest w wielu bazach. Jeśli zdefiniujemy i opiszemy te połączenia, przed użytkownikiem otwierają się zupełnie nowe możliwości korzystania z uwolnionych w internecie danych.

Na tym właśnie polega idea Linked Open Data będąca kluczowym założeniem zupełnie nowego rodzaju internetu, nazwanego siecią semantyczną lub po prostu Web 3.0. Wszystkie zasoby

są tu opisane w sposób zrozumiały dla komputera i połączone z innymi rejestrami. Dzięki temu np. mapa ewidencyjna przestaje być dla maszyny wyłącznie zbiorem poligonów, ale jest bazą granic ewidencyjnych powiązaną choćby z księgami wieczystymi czy PESEL-em. Co za chwilę udowodnimy, coraz potężniejsza moc obliczeniowa komputerów z tak zapisanych danych może zrobić niesamowity użytek.

• LOD w skrótach

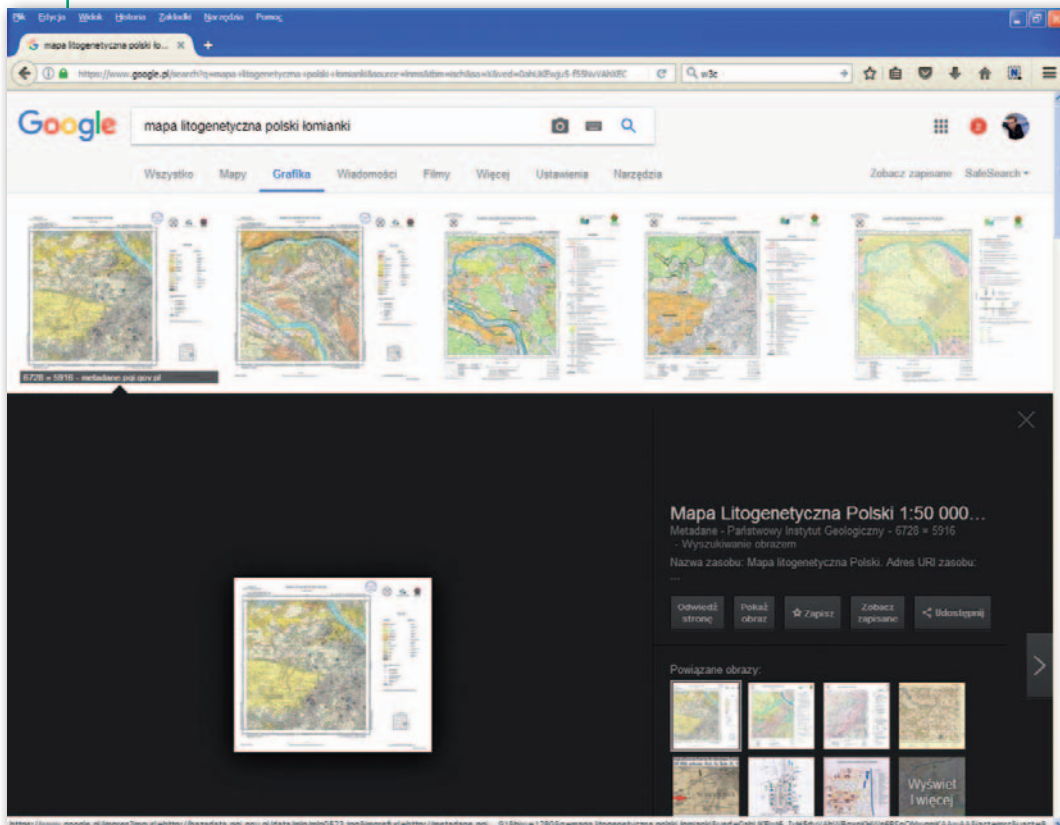
Podstawą idei Linked Open Data jest nadanie obiektom unikalnego globalnego identyfikatora, czemu służy URI (*Universal Resource Identifier*). Dzięki temu, że nadaje się mu postać adresu internetowego, ma on charakter globalny, a nie – jak w przypadku zwykłej bazy danych – tylko lokalny. Przykładowo, War-

szawa w bazie Geonames (o której więcej w dalszej części artykułu) ma URI:

<http://sws.geonames.org/756135/>.

Znaczenie tych identyfikatorów jest chyba oczywiste – chodzi o to, by komputer mógł jednoznacznie rozróżnić obiekty o tych samych lub podobnych nazwach (np. Warszawę w Polsce oraz w stanie Indiana).

Drugim poziomem wtajemniczenia są pliki RDF (*Resource Description Framework*). Zapisywane są w nich krótkie i proste zdania opisujące dany obiekt oraz jego relacje z innymi obiektami dostępnymi w sieci 3.0. Każde takie zdanie (zwane czasem trójką) składa się z podmiotu, predykatu i przedmiotu lub też z obiektu, nazwy atrybutu oraz wartości atrybutu. Podmiot to opisywany zasób, przedmiot to inny powiązany zasób identyfikowany przez URI, predykat de-



Rys. 2. Dzięki zapisowi metadanych dla arkuszy map PIG odnalezienie potrzebnego arkusza jest prostsze niż kiedykolwiek

finiuje zaś charakter relacji między podmiotem a przedmiotem. Na przykład:

```
<http://sws.geonames.org/756135/>
<http://www.geonames.org/ontology#parentCountry>
<http://sws.geonames.org/798544/>
```

oznacza, że Warszawa leży w Polsce.

Co istotne, nic nie stoi na przeszkodzie, by w pliku RDF zapisać również informacje o geometrii i położeniu geograficznym, a więc elementy typowe dla danych GIS. Współrzędne Warszawy określimy np. tak:

```
<wgs84_pos:lat>52.22977</wgs84_pos:lat><wgs84_pos:long>21.01178</wgs84_pos:long>
```

Warto wspomnieć również o standardzie RDFa, który służy do umieszczania danych RDF wewnątrz stron WWW jako atrybutów hiperlinku. Dzięki niemu otrzymujemy 2 w 1 – informację czytelną zarówno dla człowieka, jak i komputera bez zbędnego powielania zasobów.

Zgłębiając temat LOD, prędzej czy później staniemy przed pytaniem, czy w takich prostych zdaniach możemy opisać dowolną relację, nawet bardzo złożoną. Czy twórcy tego standardu przewidzieli choćby definiowanie związków między warstwami geologicznymi? Otóż sam standard RDF tu nie pomoże. Potrzebne

są sposoby, by wyjaśnić maszynie zastosowane pojęcia, a temu służą słowniki, tezauryusy oraz ontologie. W telegraficznym skrócie: słowniki są najprostszą formą takiego objaśnienia, a ontologie (które definiowane są np. w specjalnym języku OWL) – najbardziej złożoną. Kluczowym założeniem internetu 3.0. jest publikowanie tych rozwiązań w sieci, dzięki czemu raz dobrze zdefiniowana ontologia może być wykorzystywana wielokrotnie w różnych zastosowaniach.

Nieocenioną zaletą stosowania ontologii jest umożliwienie maszynie wnioskowania. Jeśli np. zdefiniujemy relacje rodzinne, a w bazie wiedzy mamy zapisane, że A jest ojcem B, a B jest ojcem C, komputer może samodzielnie wydedukować, że A jest dziadkiem C.

Wyzwaniem w sieci 3.0 jest również udostępnianie wzajemnie połączonych danych. Można oczywiście po prostu opublikować pliki RDF na serwerze, ale przecież niewiele różni się to od starego podejścia. Zdecydowanie lepszym rozwiązaniem jest przechowywanie danych w repozytorium RDF z punktem dostępowym SPARQL. SPARQL to specyficzny język zapytań do zasobów LOD, odpowiednik powszechnie stosowanego w relacyjnych bazach danych SQL-a. Jeśli tylko dysponenci danych „wystawili” odpowiednie punkty dostępowe, za pomocą jednego zapytania możemy uzys-

kać informację bazującą na zawartości kilku odrębnych, choć połączonych ze sobą baz.

• Co już mamy?

Lista zasobów dostępnych jako Linked Open Data to na razie tylko niewielki wycinek internetu, ale ich liczba rośnie w ekspresowym tempie. Przykładowo, w październiku 2007 roku w sieci można było znaleźć 2 mld „trójek RDF”, we wrześniu 2011 roku było ich już 31 mld, a w sierpniu tego roku – już 149 mld! Według szacunków ze strony stats.lod2.eu dane te przechowywane są w 9960 zbiorach, a że są one wzajemnie powiązane, relacje te można łatwo przedstawić na diagramie (rys. 1).

Jedną z kluczowych baz internetu 3.0 jest DBpedia, gdzie znajdziemy zasoby Wikipedii przetłumaczone na standardy LOD. FOAF (Friend Of A Friend) jest z kolei bazą definiującą osoby i relacje między nimi. Swoje dane w formie plików RDF oferują również Bi-

blioteka Kongresu USA, Eurostat czy Europeana (repozytorium europejskich dzieł kultury).

Jeśli chodzi o dane przestrzenne, kluczowy rejestr to wspomniany wcześniej Geonames, gdzie znajdziemy informacje o 10 milionach obiektów geograficznych z całego świata. Spore znaczenie ma także baza Linked Geodata opisująca obiekty z otwartego projektu OpenStreetMap. W przypadku urzędów za lidera należy uznać brytyjską agencję kartograficzną Ordnance Suvey, która udostępnia swoje wybrane dane w plikach RDF już od 2010 roku. Od marca br. standardy LOD wdrażane są również w szwajcarskiej infrastrukturze informacji przestrzennej.

I w Polsce urzędowe dane powoli zaczynają być udostępniane jako LOD. Na prowadzonym przez Ministerstwo Cyfryzacji portalu DanePubliczne.gov.pl znajdziemy wystawiony przez GUGiK Państwowy Rejestr Nazw Geograficznych, a także utrzymywane przez Narodowy Instytut Dziedzictwa bazy obiektów na liście UNESCO i pomników historii. Obydwa wdrożenia wykonał Wrocławski Instytut Zastosowań Informatyki i Sztucznej Inteligencji (WIZIPiSI), który jest krajowym pionierem w zakresie wykorzystania standardów LOD w zastosowaniach związanych z informacją przestrzenną. Lista projektów tej firmy jest zresztą dłuż-

sza i obejmuje zasoby: urzędów marszałkowskich województw mazowieckiego i podlaskiego (ten drugi zainteresował się LOD jako jeden z pierwszych w kraju), UM Katowice, Państwowego Instytutu Geologicznego, a także kilka innych baz CODGIK-u (poza wspomnianym PRNG). Jak wyjaśnia prezes Instytutu dr Adam Iwaniak, który naukowo tematem Linked Data zajmuje się od ponad dekady, choć bazy tych urzędów nie są jeszcze opublikowane w serwisie MC, to dostęp do nich ma każdy, choćby poprzez Google'a. Jest to możliwe dzięki opisaniu metadanych zgodnych z INSPIRE w standardzie RDFa, który jest indeksowany przez tę wyszukiwarkę. – To nasze autorskie rozwiązanie, wyróżnione zresztą jako dobra praktyka przez World Wide Web Consortium (W3C), które zajmuje się standaryzacją internetu – podkreśla prezes WIZIPiSI.

• Zabytki łączą

Za wzór uwalniania danych przestrzennych na zasadach LOD niech posłuży wspomniany już Narodowy Instytut Dziedzictwa. Kierownikiem projektu dotyczącego tego zagadnienia był Arkadiusz Kołodziej, dziś zastępca dyrektora Biura Cyfryzacji Miasta Urzędu m.st. Warszawy, gdzie zajmuje się m.in. zagadnieniami Linked Data w kontekście budowy inteligentnego miasta. Jak wspomina, bodźcem do podjęcia tamtej inicjatywy była chęć spożytkowania ogromnych nakładów pracy, jakie NID poniósł w latach 2011-2014 na budowę krajowej infrastruktury informacji przestrzennej w zakresie zabytków. – Wokół Ministerstwa Kultury i Dziedzictwa Narodowego, któremu podlega NID, podejmowanych jest wiele inicjatyw wspólnego wykorzystania danych będących w dyspozycji różnych centralnych instytucji kultury. W tym kontekście dane przestrzenne mają niezwykle ważne znaczenie – np. do opisu zasobów archiwalnych przechowywanych w Narodowym Archiwum Cyfrowym. W pewnym momencie zaczęliśmy się zastanawiać, w jaki sposób nasze dane mogą pomóc w działaniach kolegów z innych jednostek. Doszliśmy do wniosku, że rozwiązaniem będzie uspojnienie i połączenie tych zasobów, co można zrealizować właśnie dzięki LOD – mówi.

Sęk w tym, że byłby to projekt pionierski, bo na całym świecie tylko nieliczne jednostki odpowiedzialne za dziedzictwo kulturowe zdążyły mocniej wgrzyźć się w ten temat (np. English Heritage). Wsparciem okazał się WIZIPiSI. Firma dokonała konwersji czterech baz danych NID (oprócz dwóch wspomnianych wcześniej także dla zabytków archeologicznych i nieruchomości), tak aby połączyć

je z bazami Geonames, DBpedia oraz Europeana. Przygotowano także punkt dostępowy SPARQL. – Projekt ruszył pod koniec 2016 roku i udało się go zrealizować raptem w 3 miesiące, wliczając procedury formalne. Kluczowe dla zachowania tak krótkiego terminu było to, że wcześniej włożyliśmy mnóstwo wysiłku w uporządkowanie danych. Bo trzeba podkreślić, że samo uwalnianie zasobów na zasadach LOD bez należytej dbałości o ich jakość może tylko otworzyć puszkę Pandory – zaznacza Arkadiusz Kołodziej.

Trudno pominąć temat pieniędzy. – Prace kosztowały NID około 40 tys. zł. To mniej więcej tyle, ile pochłania budowa geoportalu, i to takiego na komponentach open source – podkreśla Arkadiusz Kołodziej.

Warto dodać, że dzięki realizowanemu właśnie projektowi „WIKI Wakacje” dane NID-u wzbogacą zasoby Wikipedii, ale również efekty prac wikipedystów (gromadzone w postaci „trójek”) zostaną wykorzystane podczas rozwoju portalu Zabytek.pl. – To rodzaj synergii i wspólnego wysiłku, który przynosi korzyści obu partnerom projektu i w pigułce pokazuje potencjał wykorzystania otwartych danych – mówi Arkadiusz Kołodziej.

• 6 kroków łączenia

Z punktu widzenia dysponenta danych ich otwieranie na zasadach LOD może wydawać się wrzuceniem swojej bazy do czarnej skrzynki. O to, co się konkretnie w tej skrzynce dzieje, zapytał dr. Adama Iwaniaka. Jak podkreśla, nim przystąpimy do prac technicznych, pierwszym krokiem powinno być wytypowanie zbiorów, których uwolnienie przyniesie przedsiębiorcom i społeczeństwu największe korzyści. Następnie każdy zbiór należy opisać w słownikach czy ontologiach. – Zanim się do tego zabierzemy, zawsze sprawdzamy, czy ktoś inny już wcześniej się nad tym nie pochylił. Staramy się w możliwie największym stopniu korzystać z ontologii już zdefiniowanych, wtedy stosowane pojęcia będą dla maszyn bardziej zrozumiałe – podkreśla prezes WIZIPiSI. Trzeci krok to nadanie obiektom identyfikatorów URI, a czwarty polega na konwersji danych do formatu RDF.

Kluczowy dla LOD jest krok piąty, czyli łączenie danych z różnych rejestrów. Jeśli w odrębnych bazach poszczególne obiekty mamy opisane za pomocą tych samych URI bądź innych unikalnych identyfikatorów, jest to bułka z masłem. W innym przypadku zaczynają się schody. Jak np. połączyć bazy po nazwach miejscowości, skoro Szczecin mamy zarówno na Pomorzu, jak i pod Łodzią? – Ostatnio pub-

likowanych jest mnóstwo artykułów opisujących różne metody radzenia sobie z takimi wyzwaniami. Nie ma więc jednego uniwersalnego rozwiązania tego problemu. W jednej z publikacji, której jestem współautorem (bit.ly/2fJVEOU), próbowaliśmy połączyć bazy kościołów z DBpedii i Linked Geodata. Wykorzystaliśmy tu zarówno analizy przestrzenne, jak i porównanie danych atrybutowych, przede wszystkim nazw świątyń. Pierwsza metoda była niewystarczająca, bo tam, gdzie w jednej bazie były dwa kościoły, w drugiej w tym samym miejscu widniały trzy. Drugie rozwiązanie też nie jest uniwersalne, bo bodaj w żadnym przypadku nazwy świątyń z obu rejestrów nie okazały się identyczne. Przy łączeniu baz trzeba było posługiwać się współczynnikiem prawdopodobieństwa, z jakim rekordy w obu bazach odnoszą się do tego samego obiektu w rzeczywistości, a na koniec skontrolować wynik analizy – mówi dr Adam Iwaniak.

Ostatni etap łączenia danych to ich publikacja. W projektach realizowanych przez WIZIPiSI stosuje się punkty dostępowe SPARQL, interfejsy API lub standard RDFa, czyli po prostu stronę internetową.

Oczywiście z punktu widzenia firmy realizującej projekty LOD, im cały ten proces jest bardziej automatyczny, tym lepiej. Tu niezbędne są jednak specjalistyczne narzędzia. – Przełomem w działalności naszej firmy na polu LOD było zdobycie dotacji z Narodowego Centrum Badań i Rozwoju. Dzięki niej mogliśmy stworzyć autorskie oprogramowanie Semantic Components, które dziś sprzedajemy innym podmiotom, a także sami wykorzystujemy do realizacji projektów z zakresu Linked Open Data – mówi Tomasz Berezowski wiceprezes WIZIPiSI. Prezes Instytutu zastrzega jednak, że osiągnięcie pełnej automatyzacji, choć teoretycznie możliwe, w praktyce nie zapewni rezultatów oczekiwanych przez użytkowników danych.

• Praktyczne znaczenie

Tylko na co komu te wszystkie RDF-y, URI, ontologie i zapytania SPARQL? To sztuka dla sztuki, a może użytek z tego będzie w stanie zrobić jedynie wąskie grono informatyków? No i czy tego typu rozwiązania będą miały jakiegokolwiek znaczenie dla geodezji, kartografii i GIS-u?

Otóż korzyści z publikacji geodanych na zasadach LOD można zacząć już teraz, i to bez zgłębiania specjalistycznej wiedzy informatycznej. Przykładem niech będzie wyszukanie arkusza mapy litogenetycznej dla podwarszawskich Łomianek. Większość poszukują-

cych pierwsze kroki skieruje do Google'a, a ten dotychczas odesłałby nas na stronę PIG-u, gdzie musielibyśmy wyszukać skorowidz, a następnie interesujący nas arkusz. Po pobraniu okazałoby się, że Łomianki znajdują się na dwóch arkuszach. Mniej obeznanemu internaucie cały ten proces zajmie nawet kilkanaście minut. Ale dzięki temu, że PIG opublikował metadane dla arkuszy tych map w standardzie LOD, od tego roku cała czynność jest zdecydowanie łatwiejsza – wystarczy wpisać w Google'u „mapa litogenetyczna Polski Łomianki”, kliknąć w pierwsze dwa wyniki, a następnie w odnośniki do skanów map. Czynność ta będzie szybsza, jeśli skorzystamy z wyszukiwarki grafiki (rys. 2). Takie udogodnienia są możliwe dzięki powiązaniu bazy arkuszy PIG-u z rejestrami Geonames i PRNG.

Przykład ten dobrze pokazuje, że dzięki popularyzacji LOD Google przestaje być już tylko narzędziem do wyszukiwania stron, ale również do pozyskiwania informacji.

Oczywiście większe korzyści z LOD odniesiemy wtedy, gdy opanujemy związane z tym zagadnieniem technologie, np. zapytania SPARQL. Od razu jednak uspokajam: nie taki diabeł straszny! By się o tym przekonać, wejdźmy na stronę punktu dostępowego szwajcarskiej IIP (<http://bit.ly/2uIzEuD> – rozdział *Beispielabfragen*) i skorzystajmy z jednego z dostępnych predefiniowanych zapytań. Nawet bez znajomości komend szybko zorientujemy się, jak modyfikować wybrane parametry kwerendy. Taka krótka zabawa

ze standardem SPARQL świetnie pokazuje, że popularyzacja idei LOD zrewolucjonizuje dostęp do danych przestrzennych. Wiele informacji będziemy bowiem w stanie pozyskać bez konieczności pobierania całych repozytoriów czy korzystania z GIS-owego oprogramowania.

Siła Linked Open Data – jak sama nazwa wskazuje – polega jednak przede wszystkim na łatwym łączeniu ze sobą różnych zbiorów, dzięki czemu możemy sprawnie pozyskiwać nowe informacje. Przypuśćmy, że chcemy stworzyć aplikację do wyszukiwania atrakcyjnych plenerów fotograficznych. Umożliwiłaby ona np. wskazanie gotyckiego kościoła położonego w pobliżu jeziora rynnowego. Powiązanie baz Geonames, NID i DBpedii przynajmniej teoretycznie sprawia, że taka egzotyczna analiza jest na wyciągnięcie ręki! Inny przykład: łącząc bazy EGIB i PESEL, możemy zadać pytanie: pokaż wszystkie działki należące do rodziny Jana Kowalskiego.

– Geoportal.gov.pl to niewątpliwie sukces polskiej geodezji, ale jego zawartość jest czytelna głównie dla człowieka, a nie maszyn. Jako geodeta z wykształcenia dam sobie radę z wyszukaniem np. optymalnej działki pod inwestycję, ale dla laika może być to problem. A dzięki internetowi 3.0 takie zapytanie będzie można po prostu wpisać do wyszukiwarki – mówi dr. Adam Iwaniak.

Łączenie różnych danych może być świetnym narzędziem również w pracy śledczych, którzy będą w stanie sprawnie wykryć np. skomplikowane powiązania między spółkami. Zresztą właśnie do tego

typu zastosowań powstała aplikacja Mind M.App stworzona przez WIZIPiSI, za co jej autorzy zostali docenieni nagrodą Ignite przyznaną przez Hexagon (rys. 4).

– Idea LOD świetnie wpisuje się również w popularne hasło *smart city*, bo bez otwartych danych nie ma mowy o inteligentnym mieście – mówi Arkadiusz Kołodziej. – Poza tym otwieranie danych buduje zaufanie między obywatelem i urzędem, oczywiście o ile te dane są dobrej jakości. Zwiększają także transparentność publicznych podmiotów oraz stymulują zaangażowanie mieszkańców – dodaje.

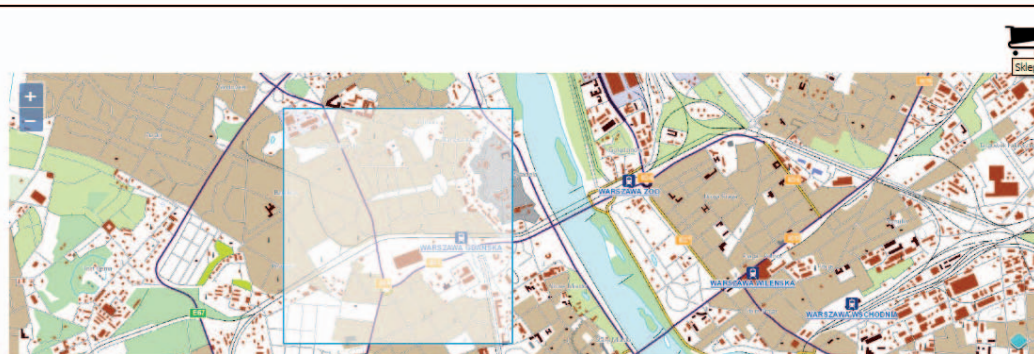
Jedną z głównych zalet wykorzystania modelu RDF jest integracja danych posiadających różny model pojęciowy. Dzięki temu, w ocenie dr Adama Iwaniaka, można łatwo rozwiązać problemy krajowej geodezji, np. rodzącego się w bólach Zintegrowanego Systemu Informacji o Nieruchomościach. – Jednym z kluczowych wyzwań związanych ze ZSIN-em jest konwersja powiatowych zasobów do centralnego repozytorium (CR). Jeśli opracujemy dedykowaną ontologię, możemy bezstratnie przekonwertować do niej dane z dowolnego powiatowego systemu informatycznego oraz wczytać je do CR, a następnie dokonać harmonizacji. Oczywiście przy tej harmonizacji zawsze były, są i będą pojawiać się problemy. Różnica polega jednak na tym, że dzięki podejściu LOD możemy powiązać dane źródłowe i zharmonizowane. W razie wątpliwości użytkownik ZSIN-u zawsze będzie mógł więc sięgnąć do tych pierwszych – wyjaśnia prezes WIZIPiSI.

Dodaje również, że LOD stanowi potencjalnie świetne pole do rozwoju małych i średnich firm z branży geoinformatycznej. – Do tego trzeba jednak całkowicie zmienić podejście do geodezji i kartografii. Urzędy powinny przestać inwestować w geoportale i mobilne aplikacje, a zamiast tego skupić się na udostępnianiu danych. Ich praktycznym wykorzystaniem niech zajmują się prywatne firmy, które same wymyślą praktyczne zastosowania i będą

Rys. 3. Po wpisaniu w wyszukiwarce „ortoobrazy Żoliborz” zostaniemy przekierowani przez Google'a do metadanych ortoobrazów przechowywanych w CODGiK-u, a tam znajduje się m.in. bezpośredni link do Portalu PZGiK, gdzie możemy nabyć to opracowanie

Ortoobrazy, 0.1, ISOK ORTO, N-34-138-B-b-4-4

8ebb8acf-1765-11e6-8129-b870f44b6730

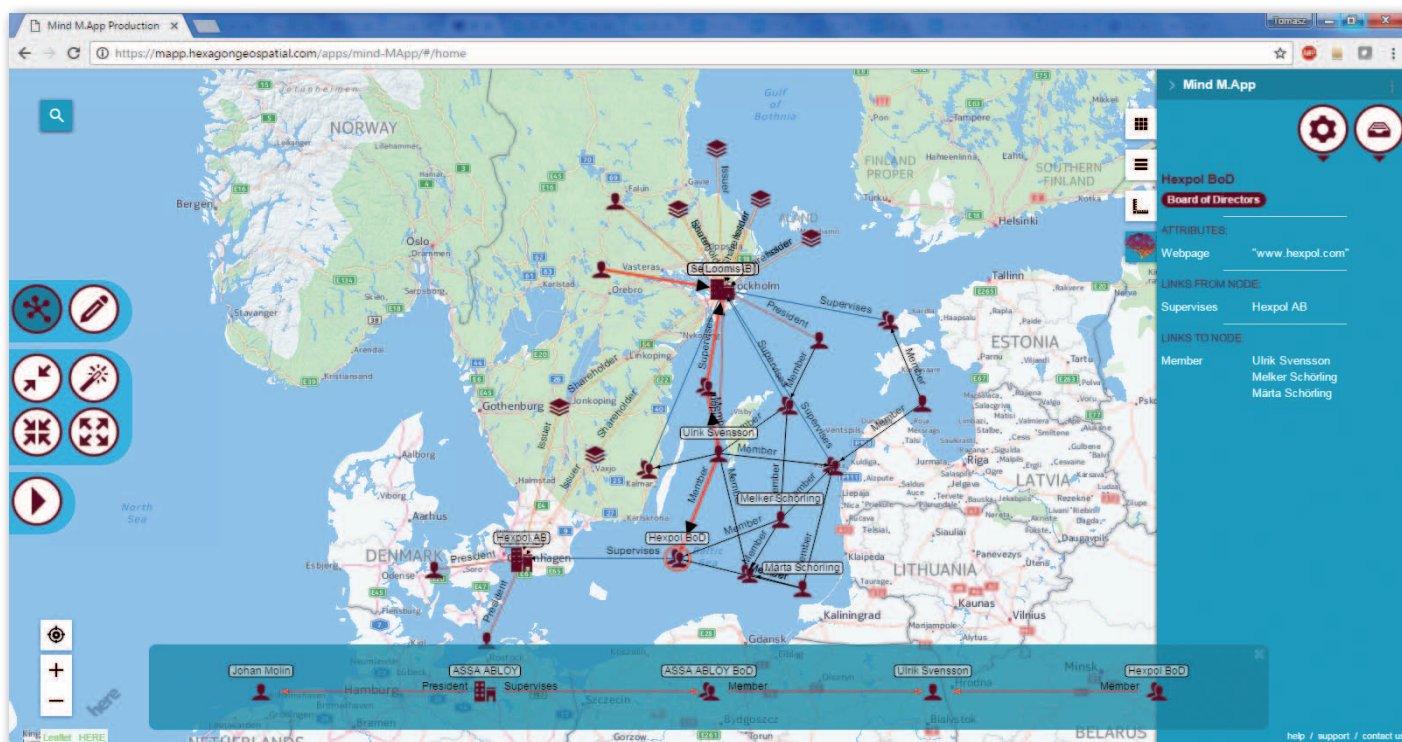


INFORMACJE PODSTAWOWE:

Tytuł:	Ortoobrazy, 0.1, ISOK ORTO, N-34-138-B-b-4-4
Kategoria:	http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/imageryBaseMapsEarthCover
Abstrakt:	Zbiór danych obejmuje metadane dla ortofotomap cyfrowych dla obszaru całego kraju. Ortofotomapy pochodzą z projektu ISOK ORTO. Zdjęcia wykonane w roku 2013. Przestrzeń barwna: RGB. Maksymalny błąd średni położenia punktu na ortofotomapie wynosi 0.30m.

SŁOWA KLUCZOWE:

Nazwa słownika podstawowego:	GEMET - INSPIRE themes, version 1.0
Słowa kluczowe:	Ortoobrazy, Zdjęcie lotnicze, Obraz terenu, Fotografia terenu, Dane referencyjne, Ortofotogram, Ortofotomapa lotnicza, Rastry, Ortofotomapa, Orthoimagery, Aerial photography, Land image, Land photography, Referential data, Raster, Aerial orthoimagery, Aerial orthophotomap



Rys. 4. Dzięki rozwiązaniom LOD stworzona przez WIZIPiSI aplikacja Mind.M.App pozwala np. analizować skomplikowane powiązania między spółkami

dą ponosić związane z tym ryzyko biznesowe – mówi.

• Dalej niż INSPIRE

Jak widać, LOD i Web 3.0 mają szansę zaferować niesamowite perspektywy rozwoju dla geodezji i kartografii. Na razie są one jednak w znacznej mierze teoretyczne. Wciąż bowiem tylko nieliczne rejestry są dostępne jako dane połączone, a jakość tych już gotowych pozostawia niekiedy wiele do życzenia. Problemem jest także brak standardów dotyczących LOD dla danych przestrzennych oraz śladowa liczba GIS-owych aplikacji gotowych do pracy na plikach RDF.

Sytuacja jednak szybko się zmienia. Jeśli chodzi o rejestry LOD – jak już wspomnieliśmy – ich liczba szybko rośnie, również w Polsce. Trwają także intensywne prace nad standardami dla połączonych geodanych. Tworzą je wspólnie Open Geospatial Consortium oraz W3C. Jednym z efektów ich pracy jest GeoSPARQL. Jak podpowiada nazwa, definiuje on zapytania SPARQL dla danych przestrzennych. Co jednak znacznie ważniejsze, rozkręcają się też pierwsze programy pilotażowe dotyczące wdrażania LOD dla zbiorów INSPIRE. Część ekspertów już wieszczy, że otwiera to drogę do budowy tzw. infrastruktury wiedzy przestrzennej.

Wdrażanie narzędzi do pracy na LOD w aplikacjach GIS na razie idzie opornie. Można wręcz odnieść wrażenie, że czołowi producenci tego typu programów

w ogóle nie dostrzegają możliwości, jakie dają połączone dane. W internecie można już jednak znaleźć narzędzia pozwalające konwertować pliki RDF do postaci strawnej dla środowiska GIS – oferuje je m.in. WIZIPiSI. Prezes Instytutu zauważa zresztą, że cały urok internetu 3.0 polega na tym, że do pracy na danych przestrzennych LOD w ogóle nie potrzebujemy specjalistycznych aplikacji – wystarczy choćby zapytania SPARQL.

Naszym krajowym problemem jest ponadto niewielkie zainteresowanie rozwiązaniami LOD. – Rozwój w tym zakresie jest imponujący. Niestety, nasze środowisko akademickie i biznesowe jest kilka lat do tyłu za resztą świata i wciąż jest zapatrzone w standardy OGC. Naprawdę ciężko mi stwierdzić, jakie są tego przyczyny – zauważa dr Adam Iwaniak.

Pozytywnym akcentem jest jednak to, że potencjał LOD dostrzegają Ministerstwo Cyfryzacji. „Z procesem otwierania danych wiążemy duże nadzieje. Szacuje się, że wartość rynku open data w Polsce w 2020 roku wyniesie ok. 1,2 mld euro. Dlatego otwieranie danych publicznych w sposób strategiczny, nastawiony na obywatela i jemu służący, to pierwszy krok do tego, żeby wykorzystywać je nie tylko na potrzeby państwa, ale także biznesu, gospodarki czy organizacji pozarządowych” – podkreśla MC. Resort zauważa ponadto, że od uruchomienia w zeszłym roku „Programu otwierania danych publicznych” systematycznie zwiększa się nie tylko liczba podmiotów udostępnia-

jących swoje zasoby, ale także stopień ich otwarcia. Na przykład już 13% zasobów na DanePubliczne.gov.pl dostępnych jest przez API. Wkrótce otwieranie danych powinno jeszcze przyspieszyć, bo MC otrzymało na ten cel z UE 23 mln zł. Resort zapowiada nawet, że rozważa wprowadzenie do ustawy o IIP zapisów wspierających ideę Linked Open Data. Na razie szczególnie nie są jednak znane.

• Zmierzch czy nadzieja?

Skoro już niedługo do przeprowadzenia analizy przestrzennej wystarczy wpisać jej treść do wyszukiwarki, a zapytania SPARQL zastąpią oprogramowanie geoprzestrzenne, na usta ciśnie się pytanie, czy nie obserwujemy zmierzchu GIS-u? – Największym grzechem środowiska geoinformatycznego jest oderwanie się od internetu i jego zasobów. GIS po prostu zamknął się w świecie swoich hermetycznych standardów. Na dziś wyzwaniem jest więc to, jak z tego wyjść. Weźmy choćby usługę WMS: to bardzo przydatne rozwiązanie, ale nie znam aplikacji spoza świata geoinformatycznego, która by z tego standardu korzystała – zauważa Tomasz Berezowski.

Zatem jeśli będziemy tkwić w starych technologiach, faktycznie możemy czuć się zagrożeni. Jeśli jednak będziemy iść z duchem czasu, przed nami otwierają się niesamowite możliwości rozwoju. Trzeba tylko mieć dobry pomysł, jak zrobić z tego użytek.

Jerzy Królikowski