

PUBLIKOWANIE DANYCH PRZESTRZENNYCH JAKO LINKED OPEN DATA*

PUBLICATION OF SPATIAL DATA AS LINKED OPEN DATA

Adam Iwaniak, Iwona Kaczmarek, Marek Strzelecki, Jaromar Łukowicz

Uniwersytet Przyrodniczy we Wrocławiu, Instytut Geodezji i Geoinformatyki

Słowa kluczowe: linked open data, dane przestrzenne, RDF

Keywords: linked open data, spatial data, RDF

Otwarte dane publiczne

Idea otwartych danych (ang. *open data*) polega na udostępnianiu danych w sposób umożliwiający ich wykorzystywanie bez żadnych ograniczeń. W obszarze geoinformacji, inicjatywy tworzenia i udostępniania otwartych danych, podejmowane są od wielu lat. Są one głównie wynikiem akcji społecznościowych, działających na zasadach całkowitego lub częściowego wolontariatu. Doskonałym przykładem jest projekt Open Street Map (<http://www.openstreetmap.org/>) rozwijający się prężnie na całym świecie, również w Polsce. W przypadku danych będących w posiadaniu administracji publicznej, sytuacja wygląda nieco inaczej. W szczególności biorąc pod uwagę istniejące uwarunkowania krajowe. Wykluczając dane, które z różnych względów nie mogą być udostępniane (np. miejsca występowania rzadkich gatunków), dane przestrzenne zgodnie z obowiązującym prawem w większości udostępnia się odpłatnie. Jakkolwiek, podejmowane w Europie działania są świadectwem na to, że idea otwartych danych, w tym danych przestrzennych, powoli staje się rzeczywistością, także w obszarze administracji.

Jednym z pierwszych kroków urzeczywistnienia pomysłu „uwolnienia” danych publicznych było opublikowanie komunikatu Komisji Europejskiej, 12 grudnia 2011 r., w którym zaprezentowano korzyści wynikające z większej otwartości danych, przedstawiono dotychczasowe osiągnięcia na tym polu, a także zaplanowano dalsze działania. Według szacunków opisanych w tym komunikacie, strategia otwartego dostępu do danych ma przynieść rocznie 40 mld euro unijnej gospodarce (Komisja Europejska, 2011a). Wyniki analiz, wykonanych na potrzeby opracowania strategii, wykazały, że istnieją znaczące trudności w zakresie zarówno wyszukiwania jak i ponownego wykorzystywania informacji z sektora publicznego.

* Artykuł sfinansowany ze środków Narodowego Centrum Nauki w ramach projektów przyznanych na podstawie decyzji DEC-2011/03/N/HS4/03819 i DEC-2012/05/B/H/HS4/04197.

W komunikacie prasowym Komisji Europejskiej (2011a) wskazano problemy związane z udostępnianiem danych z sektora informacji geograficznej. Prawie 80% respondentów wymieniło takie ograniczenia jak: wysokie opłaty, nieprzejrzyste zasady i praktyki w zakresie ponownego wykorzystywania danych, brak przejrzystości dotyczący rodzaju danych oraz podmiotów, które je przechowują, jak również porozumienia dotyczące licencji wyłącznych, których skutkiem może być zakłócanie konkurencji (Komisja Europejska, 2011a).

Obserwując działania podejmowane przez administrację publiczną w krajach członkowskich UE, można zauważyć, że ruch otwartych danych rozwija się w bardzo szybkim tempie. Coraz więcej krajów podejmuje inicjatywy w tym obszarze. Wiele z nich uruchomiło już platformy otwartych danych. Przykładem jest Wielka Brytania, która na swoim portalu <http://data.gov.uk/>, udostępnia duże ilości danych, gromadzonych przez instytucje publiczne, w tym dane przestrzenne będące w posiadaniu Ordnance Survey. Na szczęblu unijnym, funkcjonuje portal (<http://open-data.europa.eu/en/data>), stanowiący punkt dostępu do danych gromadzonych przez instytucje i inne organy Unii Europejskiej (Komisja Europejska, 2011a).

Udostępnianie danych przez agencje rządowe w popularnych formatach i na otwartych licencjach dotyczy również danych przestrzennych. Przykładem takich działań jest kanadyjski portal GeoGratis, który pozwala wyszukać i nieodpłatnie pobrać m.in. mapy rastrowe, wektorowe, dane sensoryczne oraz publikacje. Ponad 180 000 zasobów (plików do pobrania w różnych formatach np.: warstwy ESRI Shapefile, warstwy KMZ, obrazy rastrowe) zostało opublikowanych na specjalnej licencji – „Open Government Licence – Canada”, która pozwala na kopiowanie, modyfikację, publikację pobranych zasobów – również dla celów komercyjnych. Portal udostępnia narzędzia do wyszukiwania danych, które są bogato opisane z wykorzystaniem metadanych. Dzięki zastosowaniu otwartej licencji pozwalającej na komercyjne wykorzystanie danych, możliwe jest zastosowanie modelu biznesowego typu Value-Added Reselling.

Założenia Linked Open Data

Wspieranie idei otwartych danych wymaga nie tylko działań w sferze administracyjno-prawnej. Oprócz środków organizacyjnych nie mniej ważne są środki techniczne i technologiczne, a wśród nich technologie, umożliwiające efektywne wyszukiwanie i przetwarzanie danych. Rosnąca liczba informacji publikowanych w sieci WWW, wymusza poszukiwanie coraz lepszych rozwiązań służących wyszukiwaniu, udostępnianiu i integracji danych. Nadzieje upatruje się w technologiach sieci semantycznych, będących fundamentem rozwijającej się semantycznej sieci WWW (ang. *Semantic Web* lub *Web 3.0*). Według definicji proponowanej przez Berners-Lee (2001): *Semantic Web jest to rozszerzenie istniejącej sieci WWW o mechanizmy semantyczne, tak aby informacje dostępne w tej sieci były dobrze zdefiniowane i umożliwiały lepszą współpracę komputerom i ludziom*. Web 3.0 ma stanowić kolejny etap rozwoju Internetu, który wprowadzi mechanizmy inteligentnego i automatycznego przetwarzania informacji publikowanych w sieci. Obecnie większość z nich, udostępniania jest w formie czytelnej przede wszystkim dla człowieka, na przykład na stronach HTML. Przetwarzanie danych publikowanych w ten sposób przez systemy komputerowe jest w znacznym stopniu ograniczone. Często obecny etap rozwoju sieci WWW nazywany jest Internetem Dokumentów (ang. *Web of Documents*) i przeznaczony jest dla odczytu dla ludzi (ang. *human-readable*). Zamierzeniem inicjatywy Semantic Web jest stworzenie Internetu Danych

(ang. *Web of Data*), w którym udostępniane dokumenty byłyby czytelne dla maszyn (ang. *machine-readable*), miały dobrze określone znaczenie, były dostępne online i połączone z innymi zasobami.

Zbiorem wytycznych w zakresie publikowania i udostępniania danych w Semantic Web jest idea danych powiązanych, zaproponowana w 2004 przez Berners-Lee. Linked Data wykorzystuje dwie podstawowe technologie WWW: URI (Uniform Resource Identifier) i HTTP. Równie istotne jest stosowanie modelu wiązania danych RDF (Resource Description Framework), standardu wymiany danych w sieci Web, który jest rekomendowany przez W3C (2014). Zgodnie ze standardem RDF, wszystko jest zasobem, który można zidentyfikować za pomocą URI. Identyfikowana może być osoba, strona WWW, obiekt abstrakcyjny lub istniejący w rzeczywistości. URI niekoniecznie musi wskazywać na lokalizację zasobu w sieci. Zasoby w RDF reprezentowane są za pomocą zbioru stwierdzeń (ang. *statements*), zawierających trzy elementy: podmiot, orzeczenie i dopełnienie (ang. *subject, predicate, object*). Są to tzw. trójki RDF (ang. *triples*), opisujące zasoby w sposób analogiczny do tworzenia zdań, które mogą być serializowane z wykorzystaniem różnych notacji np. RDF/XML, Turtle, N-triples. Przykładowo, stwierdzenie, że Państwowa Akademia Nauk ma swoją siedzibę w Warszawie, w modelu RDF, zapisane przy pomocy notacji Turtle, przyjmuje postać:

```
@prefix dbpedia_res: < http://dbpedia.org/resource/ >  
@prefix dbpedia_prop: < http://dbpedia.org/property/ >
```

```
<dbpedia_res:Polish_Academy_of_Sciences> <dbpedia_prop:headquarters>  
<dbpedia_res:Warsaw>
```

Przykłady zasobów pochodzą z serwisu DBpedia (<http://dbpedia.org/>). Opisany powyżej zasób, posiada swoje URI o identyfikatorze http://dbpedia.org/resource/Polish_Academy_of_Sciences, oraz właściwość, która sama jest zasobem o nazwie <http://dbpedia.org/property/headquarters> (prop jest przestrzenią nazw, w której zdefiniowano i opisano znaczenie tej właściwości). Wartością właściwości jest miasto Warszawa, które także stanowi zasób o URI <http://dbpedia.org/resource/Warsaw>.

Koncepcja Linked Data opiera się na czterech podstawowych zasadach, zaproponowanych przez Tim Berners-Lee (2006):

- 1) używanie URI jako nazw dla zasobów,
- 2) używanie protokołu HTTP do lokalizacji identyfikatorów URI,
- 3) odniesienie do URI musi dostarczać użytecznej informacji o zasobie przy pomocy standardów (RDF, SPARQL Protocol and RDF Query Language),
- 4) zamieszczanie powiązań do innych URI.

Reprezentacja danych w języku RDF jest fundamentem Linked Data. Niemniej jednak, struktura trójek RDF może być bardzo rozbudowana i potrzebne są narzędzia służące na przykład do kategoryzacji elementów. Taką możliwość dają języki RDFS (RDF Schema) i OWL (Ontology Web Language), które umożliwiają między innymi definiowanie typów zasobów. Dla przedstawionego wyżej przykładu, typem zasobu reprezentującego Polską Akademię Nauk może być jednostka naukowa. Język RDFS pozwala na wyrażanie prostych konstrukcji, OWL natomiast posiada dużo większą ekspresywność. Pozwala on na definiowanie ontologii jako zbioru klas i ich właściwości (typy – część terminologiczna) oraz przykładów ich wystąpienia jako obiektów (indywidua – część opisowa). Semantyka języka OWL oparta jest na logice opisowej, co pozwala na jednoznaczną interpretację wyrażeń i aksjomatów zawartych w ontologiach (Goczyła, 2011). Ontologie w Linked Data stanowią

pewnego rodzaju standardy, które dostarczają terminologii do opisu zasobów i opisu relacji pomiędzy nimi. Przykładowo dla opisu informacji o osobach powszechnie stosuje się słownik FOAF (Friend of a friend), a do definiowania lokalizacji geograficznej ontologię Geonames (<http://www.geonames.org/ontology/documentation.html>).

Zasady Linked Data odnoszą się do wszystkich danych, niezależnie od licencji na podstawie której są udostępniane. Kombinacją otwartych danych i technologii związanych z Linked Data jest inicjatywa Linked Open Data (LOD), promowana coraz szerzej w Europie, w obszarze udostępniania danych publicznych. W 2010 r. Berners-Lee zdefiniował pięciostopniowy ranking jakości publikowanych danych (nazywany *Five stars of Open Linked Data*; tab.)

Warto zauważyć, że pierwszym elementem, jaki niezbędny jest do publikacji LOD jest otwarta licencja danych. Przy obecnych uwarunkowaniach prawnych, jedynie niewielka część danych, udostępnianych przez polską administrację publiczną, spełnia to kryterium.

Tabela. Ranking jakości danych (Berners-Lee, 2006)

*	Dostępne w internecie na otwartej licencji
**	Dostępne w postaci strukturalnej, przetwarzanej maszynowo (np. ESRI SHP zamiast jpg)
***	Dostępne w niezastereżonych standardach (WKT zamiast SHP, CSV zamiast Access)
****	Publikowane w otwartych standardach W3C RDF, URI dla identyfikacji zasobów
*****	Połączone z innymi danymi

Dane przestrzenne w Linked Open Data

Rosnące zbiory danych, publikowane zgodnie z zasadami Berners-Lee, tworzą tzw. chmurę Linked Data. Coraz więcej danych przestrzennych staje się częścią globalnej sieci. Przykładem jest przestrzenna baza trójek RDF Open Street Map, publikowanych w ramach projektu LinkedGeoData (<http://linkedgeo.org/>) obszerna baza Geonames, zawierająca ponad 8 mln obiektów geograficznych, czy też DBpedia. Najbardziej popularną ontologią służącą do wyrażania odniesienia przestrzennego jest W3C Basic Geo Vocabulary (W3C, 2006). Jest ona wykorzystywana zarówno do opisu zasobów Geonames, jak i DBpedii. Pozwala na nadanie lokalizacji geograficznej zasobom w postaci współrzędnych w układzie WGS84. Bardziej rozbudowaną propozycją jest ontologia NeoGeo. Reprezentacja obiektów przestrzennych w tym rozwiązaniu opiera się na standardzie Geography Markup Language (GML) Simple Features Profile (<http://www.ogcnetwork.net/gml-sf>). Natomiast do opisu relacji przestrzennych służy ontologia oparta na RCC8 (Region Connection Calculus).

W chwili obecnej, najbardziej popularnym rozwiązaniem wykorzystywanym do opisu danych przestrzennych w reprezentacji ontologicznej jest język GeoSPARQL, będący standardem OGC (Open Geospatial Consortium, 2011). GeoSPARQL jest jednym ze sposobów na poradzenie sobie z problemem różnorodnych i niespójnych implementacji w zakresie reprezentacji i wyszukiwania danych przestrzennych w RDF. Specyfikacja GeoSPARQL składa się z trzech głównych komponentów. Są to:

1. Terminologia do opisu obiektów przestrzennych, geometrii oraz relacji pomiędzy obiektami. Ontologia GeoSPARQL bazuje na modelu OGC Simple Features. Zawiera klasę pod-

stawową, jaką jest `geo:SpatialObject`, posiadająca dwie podklasy podrzędne `geo:Feature`, dla definicji obiektów przestrzennych oraz `geo:Geometry`, dla reprezentacji geometrii. Ontologia dostarcza szeregu klas dla reprezentacji różnych typów geometrii: linii, punktów, poligonów. Poprzez właściwości `geo:asWKT` lub `geo:asGML` dostarczana jest informacja o geometrii obiektu w reprezentacji WKT lub GML.

2. Zbiór funkcji przestrzennych do używania w zapytaniach SPARQL. GeoSPARQL dostarcza standardowego sposobu umożliwiającego zadawanie zapytań dotyczących relacji przestrzennych pomiędzy obiektami. Relacje przestrzenne mogą być wyrażone przy zastosowaniu trzech różnych słowników: OGC Simple Features, Egenhofer 9-intersection model oraz RCC8. Funkcje jakie dostarcza GeoSPARQL to np. `ogcf:intersection` (przecięcie) czy `ogcf:distance` (odległość).
3. Zbiór reguł dla transformacji zapytań.

Możliwości języka GeoSPARQL w zakresie przetwarzania danych przestrzennych są bardzo duże. Jednak pełne wykorzystanie funkcji jakie dostarcza, wymaga wykorzystania odpowiednio dedykowanych narzędzi. Liczba repozytoriów, wspierających GeoSPARQL jest w tej chwili niewielka (przykład Parliament, <http://parliament.semwebcentral.org/>). Należy mieć jednak na uwadze, że jest to standard nowy i musi minąć trochę czasu, zanim pojawią się dedykowane narzędzia i zaczną być implementowane z sukcesem w konkretnych rozwiązaniach.

Proces tworzenia danych przestrzennych w Linked Open Data

W aspekcie technicznym, proces transformacji danych przestrzennych do Linked Open Data może składać się z następujących etapów:

- 1) opracowanie ontologii do opisu danych lub wykorzystanie już istniejących,
- 2) transformacja danych do postaci RDF,
- 3) połączenie opublikowanych danych z innymi zbiorami danych,
- 4) publikacja danych.

W etapie pierwszym, dokonywane jest opracowanie lub wybór ontologii, będącej modelem dziedziny, której dane dotyczą. Na tym etapie istotne jest aby korzystać z dostępnych rozwiązań, między innymi repozytoriów ontologii, takich jak LOV (Linked Open Vocabularies, <http://lov.okfn.org/dataset/lov/index.html>) czy Swoogle, które gromadzą informacje na temat publikowanych ontologii. Jest to istotne nie tylko ze względu na uniknięcie powielania wysiłków i kosztów, ale również przyczynia się do zachowania interoperacyjności pomiędzy zbiorami danych oraz zwiększa ich wiarygodność.

Etap drugi obejmuje transformację danych do RDF. W przypadku danych przestrzennych, do tego celu wykorzystać można narzędzia ETL (ang. *Extract, transform and load*). Sama transformacja może odbywać się w sposób automatyczny, półautomatyczny lub manualny, za pomocą różnych narzędzi i metod. Przykładowo, dla potrzeb transformacji danych przechowywanych w relacyjnych bazach danych do postaci RDF służy standard W3C – język R2RML (W3C, 2012), który definiuje reguły mapowania pomiędzy modelem relacyjnym a modelem RDF. Z założenia, proces transformacji zasobów, nie posiadających ontologicznej reprezentacji (ang. *non-ontological resources*) może okazać się złożonym zadaniem,

w szczególności w przypadku, gdy podlegają mu dane o skomplikowanej strukturze, takie jak dane przestrzenne.

Trzeci etap to połączenie danych w RDF z innymi zbiorami danych. Jest to jeden z ważniejszych elementów w procesie tworzenia Linked Data. Na tym etapie konieczne jest odnalezienie powiązań pomiędzy różnymi zbiorami danych i nadanie im właściwych relacji.

Ostatni etap, jakim jest publikacja danych, obejmuje udostępnienie zarówno samego zbioru danych, metadanych, jak i dostarczenie narzędzi umożliwiających wyszukiwanie danych. Te trzy elementy realizowane są kolejno poprzez tworzenie repozytoriów trójek RDF, opracowanie metadanych dla publikowanych zasobów oraz udostępnienie punktu dostępowego do danych, np. SPARQL endpoint.

Przykłady publikacji danych przestrzennych w Linked Open Data

Potencjał Linked Data tkwi w integracji rozproszonych zasobów danych. Połączone dane, gromadzone w odrębnych repozytoriach, mogą być w łatwy sposób przeszukiwane i integrowane. Udostępniane usługi, stanowiące punkty dostępu do danych, umożliwiają zadawanie rozproszonych zapytań w języku SPARQL.

Przykład 1

Zapytanie do bazy Open Street Map (projekt Geo Linked Data) o wszystkie bankomaty w promieniu 1 km od Uniwersytetu Przyrodniczego we Wrocławiu, przedstawia się następująco:

```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc: <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>
Select *
From <http://linkedgeodata.org> {
  ?bankomat
  a lgdo:Atm ;
  rdfs:label ?nazwa ;
  geom:geometry [
    ogc:asWKT ?geometria
  ].
  Filter(bif:st_intersects (?g, bif:st_point (17.0638, 51.1114), 1)) .
}
```

Zapytanie zwraca listę punktów, które są w kategorii bankomaty (każdy posiada własne URI), nazwę bankomatu oraz jego geometrię zapisaną w formacie WKT.

Bankomat	Nazwa	Geometria
http://linkedgeodata.org/triplify/node935789248	"Bankomat BZ WBK"	"POINT(17.0623 51.1109)"^^ <http://www.openlinksw.com/schemas/virtrdf#Geometry>
http://linkedgeodata.org/triplify/node935792672	"Bankomat BZ WBK"	"POINT(17.0607 51.1082)"^^ <http://www.openlinksw.com/schemas/virtrdf#Geometry>

http://linkedgedata.org/triplify/ "Bankomat BZ WBK" node935796110	"POINT(17.0623 51.1109)"^^ <http://www.openlinksw.com/schemas/ virtrdf#Geometry>
http://linkedgedata.org/triplify/ "Bankomat BZ WBK" node935796948	"POINT(17.0575 51.111)"^^ <http://www.openlinksw.com/schemas/ virtrdf#Geometry>
http://linkedgedata.org/triplify/ "Bankomat Millenium" node935798901	"POINT(17.0642 51.1129)"^^ <http://www.openlinksw.com/schemas/ virtrdf#Geometry>
http://linkedgedata.org/triplify/ "Bankomat Cash4You" node935799446	"POINT(17.0676 51.1138)"^^ <http://www.openlinksw.com/schemas/ virtrdf#Geometry>
http://linkedgedata.org/triplify/ "Bankomat BZ WBK" node935803407	"POINT(17.0616 51.1083)"^^ <http://www.openlinksw.com/schemas/ virtrdf#Geometry>

Przykład 2

Ponieważ struktura grafu RDF, w którym publikowany jest Linked Open Data, nie jest związana sztywnym schematem, trójki zbudowane z węzłów i krawędzi mogą się dowolnie łączyć. Dlatego ważne jest, o czym wyżej pisano, aby przy określaniu typu (klasy) węzła, jak i predykatów używać powszechnie dostępne słowniki. Dla zasobów opisujących wyspecjalizowane zagadnienia, odnoszące się do dyscyplin prawnych lub społecznych usługujących się charakterystycznym językiem zawodowym, warto opracować tzw. ontologie dziedzinowe. Powinny one rozszerzać lub importować powszechnie dostępne ontologie wyższego rzędu, uzupełniając je o terminologię odrębną dla opisywanych przedmiotów.

Przykładem tego typu dziedzin są zagadnienia katastru gruntowego (ewidencji gruntów i budynków) oraz planowania przestrzennego. Poniżej przedstawiono przykład słownika definiującego pojęcia z zakresu gospodarki przestrzennej i planowania przestrzennego. Jest to ontologia łącząca różne zagadnienia: zagospodarowanie przestrzenne, relacje funkcjonalne, przestrzenne (topologiczne) oraz powiązania z innymi dziedzinami, takimi jak infrastruktura techniczna, transport czy też kataster gruntowy. Zawiera ona definicje klas obiektów oraz ich właściwości, które mogą być wykorzystane do modelowania rzeczywistych zjawisk i zapis modelu w postaci grafu RDF. Przykład ten obrazuje metody tworzenia systemu powiązań z zewnętrznymi zasobami LOD, zawartymi w zasobach GeoNames oraz sposób reprezentacji relacji topologicznych, dzięki użyciu operatorów RCC8 ze słownika NeoGeo. Prezentowany poniżej dokument zapisany został z wykorzystaniem notacji Turtle.

```
@prefix : <http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix swrl: <http://www.w3.org/2003/11/swrl#> .
@prefix swrlb: <http://www.w3.org/2003/11/swrlb#> .
@prefix owlapi: <http://www.semanticweb.org/owlapi#> .
@prefix plankb: <http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#> .
@prefix spatial: <http://geovocab.org/spatial#> .
```

```

@prefix geometry: <http://geovocab.org/geometry#> .
@prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@base <http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox> .

<http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox> rdf:type owl:Ontology ;
  rdfs:comment "Ontologia, która wprowadza definicje odniesienia przestrzennego
z ontologii NeoGeo"@pl ;
  owl:imports <http://geovocab.org/geometry> ,
               <http://geovocab.org/spatial> .

### http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#PlanningZone

### http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#LandUseUnit

:LandUseUnit rdf:type owl:Class ;
  rdfs:label "Użytek gruntowy"@pl ;
  owl:equivalentClass [ rdf:type owl:Restriction ;
                        owl:onProperty :categoryOfLandUse ;
                        owl:someValuesFrom :LandUseCategory
                      ],
                      [ rdf:type owl:Restriction ;
                        owl:onProperty :categoryOfSpatialObject ;
                        owl:hasValue :CategoryOf_CadastralLandUseUnit
                      ];
  rdfs:subClassOf :CadastralObject ;
  rdfs:comment "Użytek gruntowy"@pl .

### http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#Parcel

:Parcel rdf:type owl:Class ;
  rdfs:label "Działka geodezyjna"@pl ;
  owl:equivalentClass [ rdf:type owl:Restriction ;
                        owl:onProperty :categoryOfSpatialObject ;
                        owl:hasValue :CategoryOf_CadastralParcel
                      ];
  rdfs:subClassOf :CadastralObject ;
  rdfs:comment "Działka geodezyjna"@pl .

### http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#CadastralBuilding

:CadastralBuilding rdf:type owl:Class ;
  rdfs:label "Budynek"@pl ;
  owl:equivalentClass [ rdf:type owl:Restriction ;
                        owl:onProperty :categoryOfSpatialObject ;
                        owl:hasValue :CategoryOf_CadastralBuilding
                      ],
                      [ rdf:type owl:Restriction ;
                        owl:onProperty :categoryOfBuilding-StructureFunction ;
                        owl:someValuesFrom :Building-StructureFunctionCategory
                      ];
  rdfs:subClassOf :CadastralObject ;
  rdfs:comment "Budynek z perspektywy EGİB. Aspekt rejestrowy budynku."@pl .

###

```


http://wogis2.igig.up.wroc.pl/tbox/sdss/sdss_plankb_tbox#SpatPlanElaborationArea

```
:SpatPlanElaborationArea rdf:type owl:Class ;
  rdfs:label "Obszar opracowania planu miejscowego"@pl ;
  owl:equivalentClass [ rdf:type owl:Restriction ;
    owl:onProperty :categoryOfSpatialObject ;
    owl:hasValue :CategoryOf_SpatPlanElaborationArea
  ];
  rdfs:subClassOf :ElaborationArea .
```

```
:PlanningZone rdf:type owl:Class ;
  rdfs:label "Teren planistyczny"@pl ;
  owl:equivalentClass [ rdf:type owl:Restriction ;
    owl:onProperty :categoryOfSpatialObject ;
    owl:hasValue :CategoryOf_SpatPlanZone
  ];
  rdfs:subClassOf :SpatialPlanningObject .
```

Zgodnie z powyższym słownikiem, z danych katastralnych oraz rejestrów miejscowych planów zagospodarowania przestrzennego (w przykładzie przechowywanych w relacyjnej bazie danych z rozszerzeniem przestrzennym – PostgreSQL/PostGIS) wygenerowana została postać semantyczna opisująca te zasoby, w której relacje przestrzenne są opisane operatorami RCC8. Wygląda to następująco:

```
@prefix : <http://www.semgis.st7.eu/owl/toporel_abox.owl#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix swrl: <http://www.w3.org/2003/11/swrl#> .
@prefix swrlb: <http://www.w3.org/2003/11/swrlb#> .
@prefix plankb: <http://www.semgis.st7.eu/owl/sdss_plankb_tbox.owl#> .
```

```
@prefix spatial: <http://geovocab.org/spatial#> .
@prefix geometry: <http://geovocab.org/geometry#> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@base <http://www.semgis.st7.eu/owl/toporel_abox.owl> .
```

```
<http://www.semgis.st7.eu/owl/toporel_abox.owl> rdf:type owl:Ontology ;
  rdfs:comment „Graf relacji przestrzennych działek"@pl ;
  owl:imports <http://www.semgis.st7.eu/owl/sdss_plankb_tbox.owl> .
```

```
### http://www.semgis.st7.eu/owl/toporel_abox.owl#PL.320604_5.budynek.7b9faf74-
f730-42fb-9343-01ceff62b374.2012-11-08T09-30-29.632451_01
```

```
:PL.320604_5.budynek.7b9faf74-f730-42fb-9343-01ceff62b374.2012-11-08T09-30-
29.632451_01
  rdf:type plankb:CadastralBuilding ,
    owl:NamedIndividual ;
  rdfs:label "m2"^^xsd:string ;
  plankb:areaMeasureInSquareMeters „62.55078125"^^xsd:double ;
  rdfs:comment „Budynek m2"^^xsd:string ;
  plankb:categoryOfBuilding-StructureFunction plankb:Building-
StructureFunction_m ;
```

```

    spatial:PP :PL.320604_5.dzialka.06843cbf-cb7b-4cbd-89cf-05e27600ffcf.2012-11-08T09-30-29.632451_01 .
:PL.320604_5.dzialka.06843cbf-cb7b-4cbd-89cf-05e27600ffcf.2012-11-08T09-30-29.632451_01
    rdf:type plankb:Parcel ,
            owl:NamedIndividual ;
    rdfs:label "132/4"^^xsd:string ;
    plankb:areaMeasureInHectars "0.135797265625"^^xsd:double ;
    plankb:areaMeasureInSquareMeters "1357.97265625"^^xsd:double ;
    rdfs:comment "Działka geodezyjna nr 132/4"^^xsd:string ;
    wgs84_pos:long 14.4977115743166 ;
    wgs84_pos:lat 53.5683617424085 ;
    wgs84_pos:lat_long 53.5683617424085,14.4977115743166 ;
    gn:parentADM1 <http://sws.geonames.org/3337499/> ;
    gn:parentADM2 <http://sws.geonames.org/7530844/> ;
    gn:parentADM3 <http://sws.geonames.org/7533330/> ;
    gn:parentFeature <http://sws.geonames.org/7533330/> ;
    gn:parentCountry <http://sws.geonames.org/798544/> ;
    plankb:relationToSpatPlanArea plankb:objectCoveredBySpatialPlan ;
    spatial:PPi :PL.320604_5.budynek.7b9faf74-f730-42fb-9343-01ceff62b374.2012-11-08T09-30-29.632451_01 ,
              :PL.320604_5.budynek.ec436261-e0c3-4086-96fc-4194829877d0.2012-11-08T09-30-29.632451_01 ;
    spatial:EC :PL.320604_5.dzialka.26e65002-28b5-4401-a648-9d3e662c61aa.2012-11-08T09-30-29.632451_01 ,
              :PL.320604_5.dzialka.b0158c7b-fe57-4e0a-a57b-1e8b55d957e7.2012-11-08T09-30-29.632451_01 ,
              :PL.320604_5.dzialka.cf28f470-8f7c-4b76-b14b-b1c3b4a1f99a.2012-11-08T09-30-29.632451_01 ,
              :PL.320604_5.dzialka.f97097e5-660c-49bf-bbe3-7bd048b61584.2012-11-08T09-30-29.632451_01 ;
    spatial:O :PL.320604_5.liniaElektroEnerg.b85f1f63-556e-4eab-9f75-a29d4a419cf4.2012-11-08T11-59-03.637595_01 ,
             :PL.320604_5.liniaElektroEnerg.ee743575-9dbd-41e9-bfdd-ee2661c7b92a.2012-11-08T11-43-49.561685_01 ,
             <http://www.semgis.st7.eu/owl/toporel_abox.owl#PL.320604_5.mpzp.4c5f01c4-cfe9-42d9-8c2e-83fd94f714a0.2012-11-10T13:12:15.249454+01> ,
             :PL.320604_5.uzYTEk.97de0583-cb1b-44cb-8a10-e7b88c514ce2.2012-11-08T09-30-29.632451_01 ,
             :PL.320604_5.wodociag.1710396f-a7d6-42d7-8854-7f3c63555500.2012-11-08T11-10-16.077051_01 .

```

Dzięki referencjom do ogólnodostępnych zasobów LOD zawierających informację geograficzną oraz dzięki georeferencji z użyciem współrzędnych długość/szerokość geograficzna, każde udostępniane dane mogą być geolokalizowane. Pozwala to na tworzenie relacji przestrzennych do danych z zupełnie innych dziedzin, które również zawierają odniesienia przestrzenne (np. GeoNames, DBpedia, Open Street Map). Wspólne i referowalne odniesienia przestrzenne to podstawowy mechanizm, który umożliwia pozyskanie danych z wielu niezależnych źródeł, których wzajemne zestawienie może wynikać z bardzo specyficznych potrzeb użytkownika (np. analizy statystyczne do zaawansowanych badań społeczno-ekonomicznych). Takie wspólne odniesienie pozwala na zadawanie rozproszonych zapytań (ang.

federated queries) wzbogacających zestaw informacji stosownie do złożonych potrzeb użytkownika.

Podsumowanie

W artykule autorzy przedstawili założenia Open Data i Linked Open Data, w szczególności w zakresie udostępniania danych przestrzennych. Opisano poszczególne etapy procesu publikowania danych oraz przykłady własne danych powiązanych z dziedziny katastru i gospodarki przestrzennej.

Publikowanie danych, nie tylko przestrzennych, zgodnie z zasadami Linked Open Data wymaga podjęcia wielu działań zarówno o charakterze prawnym jak i technicznym. Wymaga także dobrej woli i współpracy pomiędzy obywatelami, przedstawicielami administracji, przedsiębiorstwami i środowiskiem naukowym. Brak świadomości wśród jednostek publicznych w zakresie możliwości, jakie dają otwarte dane z pewnością nie gwarantuje sukcesu w tym obszarze.

Można zauważyć aktywność Komisji EU i działania podejmowane przez kolejne kraje EU, które otwierają swoje zasoby danych. Jednym z podnoszonych argumentów są wyniki analiz, które potwierdzają, że model otwartych danych pozwala uzyskać lepsze rezultaty z makroekonomicznego punktu widzenia, jak również przynosi korzyści danemu organowi sektora publicznego, który je udostępnia (Komisja Europejska, 2011b). Wszelkie działania zmierzające do „uwolnienia” zasobów (Open Data) oraz jak najszerszego ich wykorzystania (Linked Open Data), pobudzają gospodarkę i kreują społeczeństwo informacyjne. Obserwując ostatnie doświadczenia innych krajów EU, można oczekiwać, że również w naszym kraju takie prace będą musiały być podjęte.

Literatura

- Berners-Lee T., Hendler J., Lassila O., 2001: The Semantic Web. *Scientific American*: 29-37.
- Berners-Lee T., 2006: Linked Data – Design Issues. Dostęp 28.02.2014 r. <http://www.w3.org/DesignIssues/LinkedData.html>
- Goczyła K., 2011: Ontologie w systemach informatycznych. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Komisja Europejska, 2011a: Komunikat prasowy, European Commission – IP/11/1524 12/12/2011, Agenda cyfrowa: Zamieniane danych rządowych w złoto. Dostęp 21.02.2014 r. http://europa.eu/rapid/press-release_IP-11-1524_pl.htm?locale=en
- Komisja Europejska, 2011b: Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiego Komitetu Ekonomiczno-społecznego i Komitetu Regionów. Otwarte dane – siła napędowa innowacji, wzrostu gospodarczego oraz przejrzystego zarządzania, KOM(2011) 882 wersja ostateczna, Bruksela, 12.12.2011 r. Dostęp 28.02.2014 r. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:PL:PDF>.
- Open Geospatial Consortium, 2011: OGC GeoSPARQL – A Geographic Query Language for RDF Data. Dostęp 28.02.2014r. Dokument 11-052r3. <http://www.opengeospatial.org/standards/requests/80>.
- Salas J.M., Harth A., Norton B., Vilches L.M., León A.D., Goodwin J., Stadler C., Anand S., Harries D., 2012: Neo-Geo Vocabulary: Defining a shared RDF representation for GeoData. Public draft, NeoGeo. Dostęp 28.02.2014 r. <http://geovocab.org/doc/neogeo.html>
- W3C, 2006: W3C Semantic Web Interest Group: Basic Geo (WGS84 lat/long) Vocabulary, 2006. Dostęp 28.02.2014r. <http://www.w3.org/2003/01/geo/>

W3C, 2012: R2RML: RDB to RDF Mapping Language, R2RML: RDB to RDF Mapping Language. Dostęp 28.02.2014r. <http://www.w3.org/TR/r2rml/>

W3C, 2014: RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation 25 February 2014. Dostęp 28.02.2014 r. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

Źródła internetowe

<http://www.geonames.org/>

<http://linkedgeodata.org/>

<http://www.ogcnetwork.net/gml-sf>

<http://parliament.semwebcentral.org/>

Linked Open Vocabularies (LOV). <http://lov.okfn.org/dataset/lov/index.html>

DBpedia. <http://dbpedia.org/>

Open Street Map. <http://www.openstreetmap.org/>

Streszczenie

Udostępnianie informacji publicznej jest tematem wiodącym od wielu lat na całym świecie. Podejmowane na świecie inicjatywy, służące udostępnianiu danych rządowych, w tym danych przestrzennych, wskazują na istotne korzyści wynikające z otwarcia publicznych zasobów. Najlepszym dowodem jest fakt, iż coraz więcej krajów rozpoczyna działania zmierzające do „uwolnienia” swoich zasobów. Open Data, czy też w przypadku danych rządowych Open Government Data to inicjatywy mające na celu udostępnienie danych dla społeczeństwa w postaci, która umożliwi ich swobodne wykorzystanie bez ograniczeń. Oznacza to, że dane udostępniane są na otwartych licencjach i w otwartych standardach. Pozwala to potencjalnym użytkownikom na kreowanie nowych usług i produktów, np. w postaci użytecznych aplikacji czy serwisów bazujących na dostępnych danych przestrzennych, co jest motorem pobudzającym biznes i gospodarkę. W dobie serwisów społecznościowych, jest to również metoda na weryfikację danych rządowych i podniesienie ich jakości. Istotą udostępniania danych Open Data jest nie tylko stosowanie otwartych standardów do ich publikacji. Niemniej ważną kwestią jest publikacja danych, które mogłyby być nie tylko czytelne dla człowieka (ang. human-readable), ale również przetwarzane przez maszyny (ang. machine-readable). Podejście to opiera się na wykorzystaniu technologii rozwijającego się Semantic Web i określane jest mianem Linked Data. Publikowanie danych otwartych w technologii Linked Data zyskało nazwę Linked Open Data (LOD).

Publikacja danych przestrzennych w chmurze LOD wiąże się ze spełnieniem kilku warunków, uznanych jako ogólne wytyczne w zakresie tworzenia Linked Open Data. Po pierwsze, dane muszą zostać udostępnione na otwartej licencji, umożliwiającej ich dowolne wykorzystanie. Po drugie, dane muszą mieć formę ustrukturyzowaną. Po trzecie muszą być udostępniane w otwartych standardach, np. WKT zamiast Shapefile. Po czwarte, reprezentacja danych musi być możliwa w modelu RDF (ang. Resource Description Framework) oraz każdy obiekt musi mieć nadany unikalny identyfikator URI. Ostatnią kwestią jest łączenie udostępnianych zasobów z innymi poprzez tworzenie relacji pomiędzy nimi. Grupa technologii w zakresie Linked Open Data obejmuje standard do reprezentacji danych RDF, język zapytań SPARQL, globalny identyfikator URI oraz języki do modelowania ontologii OWL i RDFS. Wykorzystanie ww. technologii stanowi nowe podejście w zakresie reprezentacji, pozyskiwania, integracji i wyszukiwania danych przestrzennych. Pomimo, niedojrzałych jeszcze rozwiązań technicznych w tym zakresie, inicjatywa Linked Open Data wkracza wielkimi krokami w obszar geoinformacji. Podejmowane działania w wielu krajach na świecie, wskazują, że udostępnianie rządowych danych przestrzennych jako Open Data jest potrzebne i niewątpliwie przynosi korzyści ekonomiczne. Istnieje jednak silna potrzeba zapoczątkowania wielkiej zmiany, przede wszystkim w sposobie myślenia, co w rzeczywistości może stanowić największą przeszkodę i trudność.

Abstract

For many years, issues associated with sharing of public information have been one of the major topics. The initiatives undertaken all over the world, with the purpose of providing governmental data – including spatial data – have proven that there are significant benefits from the opening of these resources. The fact that more and more governments start to make efforts to “liberate” their data is the best proof. Open Data, or in this case more precisely – Open Government Data are the initiatives for sharing data with the society in such forms that it can be used freely and without restrictions. This means that data is published on open licenses and with the use of open standards. It enables potential users to create new services and products, for example useful applications or services based on published spatial data, which can provide a stimulus for economy and business. In the age of social media, it is also a way for data verification, which can improve its quality. But the essence of publishing Open Data is not only the use of open standards, but also providing data which can be not only human readable but also machine readable. This approach is based on and strictly connected with Semantic Web technologies and it is called Linked Data. Publishing open data by means of Linked Data has gained the name of Linked Open Data (LOD).

Publishing spatial data in LOD cloud involves meeting few requirements, known as the principles of creating Linked Open Data. Firstly, the data must be published on open license, ensuring that the data can be used without restrictions. Secondly, the data must be in structured representations. Thirdly, it must be published with the use of open standards, for example WKT instead of Shapefile. Moreover, it must be possible to represent the data with the use of RDF (Resource Description Framework) model and every entity within the data set must have a unique URI identifier. Last but not least, it is important to link newly published resources with the other by creating relations between them. Linked Open Data technologies consist of data representation model RDF, query language SPARQL, URI identifiers and ontology modeling languages OWL and RDFS. With the use of these technologies a new approach to spatial data representation, acquisition, integration and discovery emerges. Although the technologies connecting this approach with spatial data are still immature, the Linked Open Data enters the field of geomatics. The actions undertaken in many countries around the world prove that publishing and providing governmental spatial data as Open Data is crucial and is undoubtedly beneficial for the economy. However, there is a strong need to initiate this great change, especially in the way of thinking, which appears to be the biggest obstacle and difficulty.

dr inż. Adam Iwaniak
adam.iwaniak@up.wroc.pl

mgr inż. Iwona Kaczmarek
iwona.kaczmarek@up.wroc.pl

mgr inż. Marek Strzelecki
marek.strzelecki@kon-dor.pl

mgr inż. Jaromar Łukowicz
jaromar.lukowicz@struktura.eu