

## ROLA TEZAURUSA W KSZTAŁTOWANIU INTEROPERACYJNOŚCI SEMANTYCZNEJ

### THE ROLE OF THESAURUS IN CREATING SEMANTIC INTEROPERABILITY

Iwona Kaczmarek, Adam Iwaniak

Uniwersytet Przyrodniczy we Wrocławiu  
Kon-Dor GIS Consulting, Wrocław

**Słowa kluczowe:** interoperacyjność semantyczna, tezaurus, SDI, SKOS

Keywords: semantic interoperability, thesaurus, SDI, SKOS

### Interoperacyjność w SDI

Koncepcja oraz dalszy rozwój infrastruktury danych przestrzennych (ang. *Spatial Data Infrastructure* – SDI) są konsekwencją rozwoju internetu i WWW. Paradygmat w jakim budowana jest SDI oparty jest na architekturze zorientowanej na usługi SOA (ang. *Service Oriented Architecture*). W aspekcie rozwiązań sieciowych pojawia się zatem problem interoperacyjności, zarówno w zakresie wymiany danych jak i komunikacji usług.

Pojęcie interoperacyjności nie jest pojęciem nowym i utożsamiane jest głównie z systemami informatycznymi. Najczęściej spotykana i ogólna definicja interoperacyjności określa ją jako *zdolność dwóch lub większej liczby komponentów do wymiany informacji, rozumienia jej oraz wykorzystania* (IEEE Standard, 1990). Osiągnięcie interoperacyjności umożliwi wspólną komunikację pomiędzy systemami oraz wymianę i wykorzystanie tych samych danych w różnych systemach. Według ISO/IEC 2382-01, Information Technology Vocabulary, Fundamental Terms (2003), interoperacyjność oznacza zdolność różnych elementów funkcjonalnych systemów informatycznych do komunikacji, uruchamiania programów lub przesyłania danych pomiędzy nimi w sposób nie wymagający od ich od użytkownika żadnej wiedzy lub wymagający od niego wiedzy minimalnej na temat unikalnych właściwości tych elementów. W Europejskich Ramach Interoperacyjności – EIF opublikowanych przez IDABC (2004) wyróżnia się trzy poziomy interoperacyjności:

1. *techniczny* – łączący systemy komputerowe i usługi,
2. *organizacyjny* – definiujący cele administracyjne, modelujący procesy administracyjne i indukujący współpracę administracji,
3. *semantyczny* – zapewniający, że dokładne znaczenie wymienianej informacji jest zrozumiałe przez inną aplikację, która pierwotnie nie została opracowana do tego celu.

Analogiczna kategoryzacja interoperacyjności, ale w odniesieniu do infrastruktury danych przestrzennych przedstawiona została przez J. Gaździckiego (2002):

1. *interoperacyjność techniczna obejmuje aspekt systemowy (urządzenia, protokoły transmisji, systemy operacyjne) oraz aspekt syntaktyczny (języki, formaty). Podstawowe znaczenie ma koncepcja SOA (Service Oriented Architecture), a zwłaszcza technologia Web Services.*
2. *interoperacyjność semantyczna dotyczy właściwego, jednoznacznego rozumienia wymienianej i upowszechnianej informacji przez wszystkich jej użytkowników.*
3. *interoperacyjność organizacyjna uwarunkowana przepisami prawnymi, strukturami i procedurami organizacyjnymi, czynnikami ekonomicznymi i kadrowymi.*

Uszczegółowiając, interoperacyjność techniczna dotyczy zagadnień związanych z zapewnieniem odpowiednich środków technicznych, takich jak standardy wymiany i prezentacji danych, standardy komunikacji usług, bezpieczeństwo usług itp. Do ich rozwoju istotnie przyczynił rozwój Internetu, dzięki czemu są one szeroko wykorzystywane. W obszarze SDI kluczową rolę odgrywają takie organizacje jak OGC (ang. *Open Geospatial Consortium*) czy ISO (ang. *International Organization for Standardization*). Standardy, stanowiące produkty ich prac, wykorzystywane i stosowane są przez Komisję Europejską w ramach budowy Europejskiej Infrastruktury Informacji Przestrzennej.

Interoperacyjność organizacyjna związana jest z wszelkimi działaniami o charakterze organizacyjnym i prawnym, zmierzającymi do zapewnienia współpracy wszystkich aktorów uczestniczących w procesie tworzenia systemu. W zakresie budowy SDI obejmuje ona wszystkie jednostki współtworzące infrastrukturę – administrację publiczną i sektor prywatny zainteresowany tworzeniem i udostępnianiem danych przestrzennych.

Działania w obszarze zapewnienia interoperacyjności technicznej i organizacyjnej w budowie SDI są podejmowane od dłuższego czasu i wdrażane w praktyce. Dużo bardziej złożone są zagadnienia związane z osiągnięciem tzw. interoperacyjności semantycznej, które ciągle pozostają w początkowej fazie badań. Samo pojęcie jakim jest semantyka w kontekście językoznawczym oznacza naukę o znaczeniu wyrazów, natomiast w aspekcie logiki, określa relacje pomiędzy znakami a rzeczywistością, do której te znaki się odnoszą. W logice semantyka stanowi odrębną dziedzinę, w której wyróżnia się dwa główne kierunki: semantykę empiryczną, koncentrującą się na językach naturalnych oraz semantykę teoretyczną zajmującą się językami formalnymi.

W infrastrukturze informacji przestrzennej, interoperacyjność semantyczną można rozpatrywać w kilku aspektach. Jeden z nich związany jest z komunikacją usług sieciowych, kluczowych dla funkcjonowania infrastruktury. Syntaktycznego opisu interfejsu usług dostarcza WSDL (ang. *Web Services Description Language*), opisujący parametry usługi – sposób jej wywołania, dane wejściowe oraz strukturę komunikatu wyjściowego. UDDI (ang. *Universal Description, Discovery, and Integration*) stanowi rejestr zawierający opis usług. Usługi, które zawarte są w rejestrze mogą być wyszukane. Metadane zapisane w rejestrze są przeznaczone przede wszystkim do odczytu i interpretacji przez człowieka, zatem tylko w pewnym stopniu stanowią sformalizowany sposób odnajdywania i uruchamiania właściwych usług. Usługi pochodzące od różnych dostawców, oferujące te same funkcje, mogą być różnie opisane. W celu organizowania usług w oparciu o ich znaczenie, a nie tylko słowa kluczowe znajdujące się w ich tekstowym opisie, należy uzupełnić opisy o semantykę, która pozwoli aplikacjom rozpoznać, czy dane wyrażenia są tożsame bądź zrozumieć, jakie są relacje pomiędzy nimi.

Drugi aspekt związany jest z semantyką danych przestrzennych wykorzystywanych w ramach SDI. W tym kontekście interoperacyjność semantyczną należałoby rozumieć jako dążenie do stanu, w którym znaczenie informacji przestrzennej jest jednoznacznie zrozumiałe. Dzięki niej możliwa jest między innymi automatyczna integracja i przetwarzanie danych przestrzennych w skali globalnej.

Interoperacyjność semantyczna stanowi kluczowy element dla dalszego rozwoju infrastruktury informacji przestrzennej i jej przejścia w infrastrukturę wiedzy przestrzennej. Trudności w osiągnięciu interoperacyjności semantycznej związane są nie tylko z kwestiami technicznymi, ale barierą stanowić może również brak zrozumienia, iż standaryzacja semantyczna jest tak samo ważna jak standaryzacja syntaktyczna.

## Sieci semantyczne

Aspekty semantycznej interoperacyjności w dziedzinie systemów informatycznych są przedmiotem badań od wielu lat. Istotną rolę w tym obszarze odgrywa XML, standard *de facto* reprezentowania informacji. Jednakże sam XML opisuje dane jedynie w sposób strukturalny, pozbawiony semantyki. Dla jej osiągnięcia konieczne staje się stosowanie dodatkowych specyfikacji, takich jak RDF (ang. *Resource Description Framework Language*) czy OWL (ang. *Web Ontology Language*), stanowiących wyniki prac organizacji standaryzacyjnej W3C (ang. *World Wide Web Consortium*) nad technologiami sieci semantycznych. Prace nad ww. standardami są działaniami podejmowanymi w celu osiągnięcia wizji *Semantic Web* (SW). Głównym założeniem SW jest publikowanie informacji w Internecie w sposób, umożliwiający ich rozumienie i przetwarzanie przez maszyny. Oznacza to, że dane pochodzące z wielu źródeł stają się przetwarzalne w sposób automatyczny, przez co mogą zostać udostępnione inteligentnym wyszukiwarkom i aplikacjom, mogącym wykorzystać je w różnych zastosowaniach. W celu osiągnięcia interoperacyjności semantycznej sieci semantyczne odgrywają zatem podstawową rolę. Dodatkowo dostarczają wiele innych możliwości w zakresie integracji i przetwarzania danych, które w dotychczasowym WWW są trudne do spełnienia. Kluczowe znaczenie odgrywa wymieniony już wcześniej standard RDF, pozwalający na opis zasobów w postaci grafu składającego się z trójek w postaci: {podmiot, predykat, obiekt}.

Dużo większe możliwości w zakresie integracji i wymiany danych dostarczają ontologie, które pozwalają na stworzenie modelu danej dziedziny i opis wybranego fragmentu rzeczywistości. Dla opracowania ontologii wykorzystywana jest często logika opisowa, która dodatkowo wprowadza mechanizmy wnioskowania, umożliwiające wysoki poziom analizy i przetwarzania danych. Klasyczne rozumienie ontologii w informatyce, zdefiniowane zostało przez T. Grubera (1993): *Ontologia jest formalną, jednoznaczną specyfikacją dzielonej (wspólnej) konceptualizacji, [...] jest opisem pojęć i relacji zachodzących między nimi. [...] Praktycznie ontologia jest zbiorem definicji formalnego słownika*. Ontologia opisuje pewien model rzeczywistości, wykorzystując do tego celu konceptualizację oraz hierarchizację. Według Fensela (2004) konceptualizację można rozumieć jako abstrakcyjny model dowolnego zjawiska świata rzeczywistego, który rozpoznaje koncepty (abstrakcyjne idea, wyobrażenie lub byt) tego zjawiska.

Istotny wpływ na rozwój wizji *Semantic Web* mają nowe formy systemów organizacji wiedzy (SOW, ang. *Knowledge Organisation Systems*) tj. sieciowe systemy organizacji. Zalicza się do nich narzędzia porządkowania wiedzy o różnym stopniu formalizacji, między

innymi słownictwo kontrolowane, stanowiące zbiór terminów, które są powszechnie akceptowane, jednoznacznie zdefiniowane oraz zarządzane. Słownictwo kontrolowane, w zależności od struktury, przyjmuje formę list, słowników, taksonomii czy tezaursów, służących do indeksowania i kategoryzowania zasobów. W systemach informatycznych spełnia ono przede wszystkim rolę wspomagania użytkownika w odnalezieniu interesujących go danych i ich poprawnej interpretacji.

Tworzenie słownictwa kontrolowanego w obszarze SDI ma szczególne znaczenie w przypadku tworzenia metadanych dla zbiorów danych przestrzennych. Ma ono za zadanie ograniczyć dowolność w opisie zasobu, przez co zwiększyć możliwości wyszukiwania informacji. Stosowanie różnych pojęć posiadających to samo znaczenie lub tych samych pojęć posiadających różne znaczenie w zależności od kontekstu powoduje trudności dla użytkownika w odnalezieniu i interpretacji danych. Dlatego, przydatnym narzędziem stają się struktury w postaci taksonomii czy też bardziej rozbudowane – w postaci tezaursów.

## Definicja tezaury

Pojęcie tezaury pochodzi ze środowiska bibliotekarskiego, w którym takie struktury organizacji informacji tworzone były w celu tematycznej klasyfikacji zasobów (książek, dokumentów). Obecnie, wraz z rozwojem technologii internetowych, ich cyfrowe wersje są stosowane w repozytoriach danych (systemach zarządzania, bibliotekach danych) w celach zarówno klasyfikacji jak i wyszukiwania.

Klasyczna definicja tezaury określa go jako słownik pojęciowy, zawierający powiązane hierarchicznie i semantycznie terminy, zwane również deskryptorami. Często definiowany jest również jako słownik tzw. języka informacyjno-wyszukiwawczego, mający za zadanie kontrolowanie słownictwa, służącego do wyszukiwania i indeksowania informacji. Cel ten realizowany jest przez wykluczenie takich cech języka naturalnego – z punktu widzenia wyszukiwania informacji – jak synonimia i wieloznaczność. Zestaw związków najczęściej występujących w teaurusie to relacje hierarchiczne, równoważnościowe i skojarzeniowe. Zgodnie ze standardami dotyczącymi budowy tezaursów, związki te oznaczane są jako BT (*broader term*), NT (*narrower term*), U/UF (*use/used for*), RT (*related term*) oraz w niektórych normach rozszerzalne o bardziej szczegółowe specyfikacje powiązań, tj. BTG (*broader term generic*).

Zatem, w tradycyjnej funkcji, tezaurus służy głównie jako narzędzie indeksowania i wyszukiwania informacji. Indeksowanie polega na identyfikacji pojęć jakie występują w danym zasobie i przedstawieniu ich za pomocą terminów (deskryptorów) pochodzących z tezaury. Wyszukiwanie w obszernych zbiorach danych z wykorzystaniem tych samych deskryptorów, które zostały użyte przy indeksowaniu, umożliwia odnalezienie wyłącznie tych informacji, które są istotne dla celu wyszukiwania.

Istnieje wiele norm i standardów, określających sposób i zasady tworzenia tezaury. Najistotniejsze z nich to:

- ISO 2788: Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri,
- ISO 5964: Documentation – Guidelines for the Establishment and Development of Multilingual Thesauri,

- ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies,
- BS 8723: Structured vocabularies for information retrieval – Guide:
  - Exchange formats and protocols for interoperability (DD 8723-5:2008). British Standard (2008),
  - Vocabularies other than thesauri (BS 8723-3:2007). British Standard (2007),
  - Interoperability between vocabularies (BS 8723-4:2007). British Standard (2007),
  - Definitions, symbols and abbreviations (BS 8723-1:2005). British Standard (2005),
  - Thesauri (BS 8723-2:2005). British Standard (2005).

Na szczeblu krajowym obowiązuje polska norma PN-92/N-09018: *Tezaurusz jednojęzyczny. Zasady tworzenia, forma i struktura*. Brak natomiast polskiej normy dotyczącej tworzenia tezaurusów wielojęzycznych.

Obecnie trwają przygotowania nad nową normą ISO 25964: *Information and documentation – Thesauri and interoperability with other vocabularies*, która będzie składała się z dwóch części – *Thesaurus for information retrieval* oraz *Interoperability with other vocabularies*. Draft części pierwszej – ISO/DIS 25964-1 został udostępniony do publicznego opiniowania pod koniec lutego 2010r. Norma ta jest o tyle istotna, iż traktuje tezaurus jako system działający w środowisku sieciowym i porusza kwestie związane z interoperacyjnością wielu systemów organizacji wiedzy. Ma za zadanie zastąpić dotychczasowe normy ISO 2788 i ISO 5964, jak również BS 8723.

## Tezaurus w dobie Web 3.0

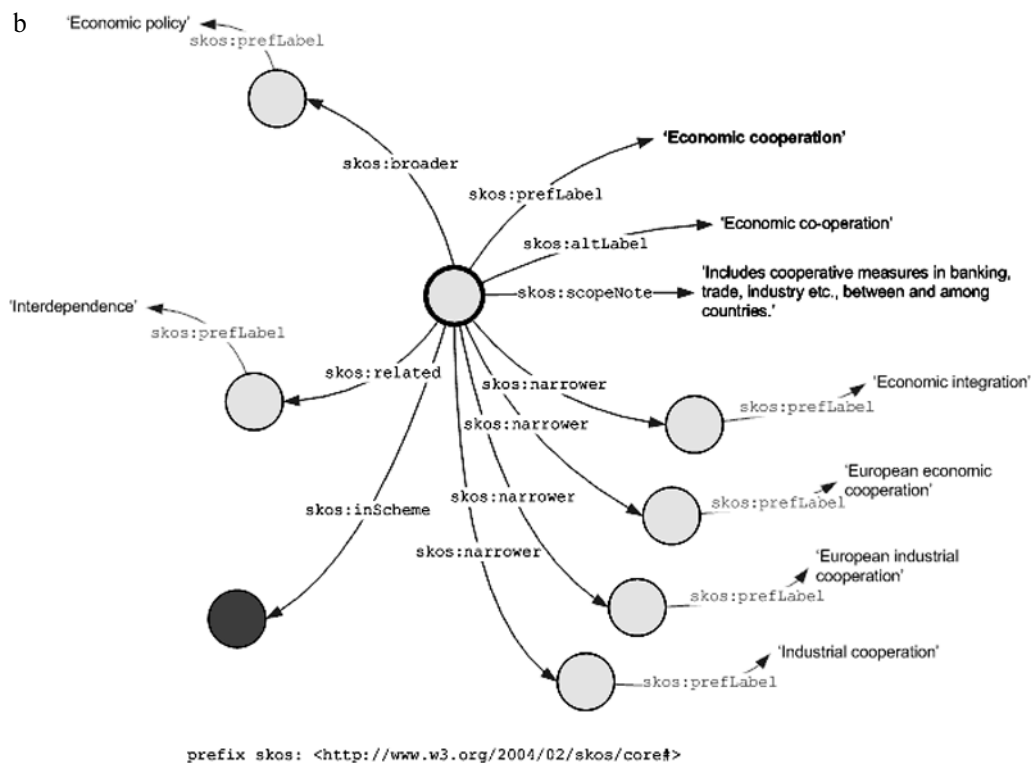
Standardem *de facto* w nowoczesnych systemach organizacji wiedzy jest specyfikacja SKOS (ang. *Simple Knowledge Organization System*) stanowiąca rekomendację W3C. Podstawowym elementem SKOS jest pojęcie (ang. *concept*), które posiada unikalne URI (ang. *Uniform Resource Identifier*). SKOS wykorzystuje składnię RDF do opisu danych. Specyfikacja SKOS określa zestaw klas i własności jakie służą do opisywania pojęć. Są to m.in.:

- klasy: *CollectableProperty*, *CollectionConcept*, *ConceptScheme*, *OrderedCollection*
- własności: *altLabel*, *altSymbol*, *broader*, *changeNote*, *definition*, *editorialNote*, *example*, *hasTopConcept*, *hiddenLabel*, *historyNote*, *inScheme*, *isPrimarySubjectOf*, *isSubjectOf*, *member*, *memberList*, *narrower*, *note*, *prefLabel*, *prefSymbol*, *primarySubject*, *related*, *scopeNote*, *semanticRelation*, *subject*, *subjectIndicator*, *symbol*.

Wiele z publikowanych tezaurusów udostępnianych jest w standardzie SKOS. Rysunek 1 przedstawia fragment pochodzący z tezaurusu UKAT (<http://www.ukat.org.uk/>) w tradycyjnym zapisie oraz w notacji SKOS.

Innym przykładem tezaurusu, który dostępny jest w SKOS jest GEMET (*General Multilingual Environmental Thesaurus*) (<http://www.eionet.europa.eu/gemet>). Posiada on szczególne znaczenie w obszarze SDI, w zakresie sporządzania metadanych. Zgodnie z Rozporządzeniem w zakresie metadanych (Rozporządzenie, 2008): *Jeżeli zasób jest zbiorem danych przestrzennych lub serią zbiorów danych przestrzennych, należy wykorzystać co najmniej jedno słowo kluczowe, pozyskane z ogólnego wielojęzycznego tezaurusu środowiskowego (GEMET), opisujące odpowiedni temat danych przestrzennych określony w załącznikach I, II lub III do dyrektywy 2007/2/WE.*

- a
- Term: Economic cooperation
- UF - Used For (Używaj) :  
Economic co-operation
- BT - Broader terms (Terminy szersze):  
Economic policy
- NT - Narrower terms (Terminy węższe):  
Economic integration  
European economic cooperation  
European industrial cooperation  
Industrial cooperation
- RT - Related terms (Terminy skojarzone):  
Interdependence
- SN - Scope Note (Zakres użycia terminu):  
Includes cooperative measures in banking, trade, industry etc., between and among countries.



Rys. 1. Fragment tezaury UKAT: a – w standardzie SKOS, b – w tradycyjnym zapisie  
(źródło: <http://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050517>)

GEMET jest tezauresem obejmującym ogólną terminologię z zakresu środowiska (rys. 2). Prace nad jego tworzeniem podjęte były z inicjatywy Europejskiej Agencji Środowiska w roku 1996. Struktura GEMET jest rozszerzeniem specyfikacji SKOS o dodatkowe elementy nie występujące w większości tezaursów. Występują tu supergrupy i grupy, w których

The screenshot shows the GEMET thesaurus interface. On the left, there is a 'Local navigation' menu with links like 'User directory', 'Roles', 'NFP/Eionet IG', etc. Below it is a 'Find a person' search box and 'Account services' like 'lost my password'. The main content area displays the entry for 'geodezja'. At the top, there are navigation links: 'Thematic Listings', 'INSPIRE Spatial Data Themes', 'Alphabetic Listings', 'Hierarchical Listings', and 'Search Thesaurus'. Below these is a language selection bar with 'zh-CN' selected. The entry for 'geodezja' includes a definition: 'nauka o pomiarach Ziemi, zajmująca się wyznaczaniem kształtu i rozmiarów globu ziemskiego lub jego części, sporządzaniem map i planów geodezyjnych, pomiarami gruntów i obiektów dla celów gospodarczych, technicznych, wojskowych itp.'. It also lists 'broader terms' (geografia), 'scope note' (scope note is not available), 'Groups' (Badania naukowe, nauka), 'Themes' (badania naukowe, procesy naturalne), and 'Other relations' (AGROVOC: Geodesy, EuroVoc: geodesy, Wikipedia article: Geodesy, UMLTHES: Geodásie). On the right, there is a list of translations for 'geodezja' in various languages: Arabic, Basque, Bulgarian, Catalan, Chinese, Czech, Danish, Dutch, English, English (US), Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Romanian, and Russian.

Rys. 2. Fragment tezaury GEMET (źródło: <http://www.eionet.europa.eu/>)

kolekcjonowane są pojęcia. Dodatkowo pojęcia mogą być powiązane z tematami, dostarczając innej strukturalizacji deskryptorów. W chwili obecnej GEMET stanowi zbiór ponad 6000 terminów dostępnych w 27 językach.

## Przykłady tezaurów

Poniżej przedstawione zostały przykłady tezaurów, których autorzy są współwykonalcami. Są one wstępnym etapem dla porządkowania wiedzy z danej dziedziny oraz niezbędne dla opracowania metadanych w sposób uporządkowany i spójny.

### Tezaurus dla edytora metadanych

Głównym zamierzeniem przy budowie tezaury jest potrzeba uporządkowania pojęć występujących w danej dziedzinie. Tezaurus może być również budowany dla dedykowanej aplikacji czy systemu, przykładowo zbiorów bibliotecznych. Dla potrzeb edytora metadanych dla zbiorów danych przestrzennych stworzony został tezaurus, którego autorami jest Instytut Geodezji i Kartografii w Warszawie oraz firma Kon-Dor GIS Consulting. Stanowić ma on źródło słów kluczowych, które wykorzystywane będą przy opisie metadanymi zbiorów państwowego zasobu geodezyjnego i kartograficznego (PZGiK). W przypadku budowy tego typu tezaury istotny jest zatem wybór słów kluczowych, które miałyby opisywać poszczególne zbiory danych. Nie mniej ważną rolę odgrywa także opracowanie struktury tezaury. Racjonalne wydaje się wykorzystanie konstrukcji tezaury GEMET, który wykorzystywany jest obligatoryjnie przy sporządzaniu metadanych.

W opracowanym tezaurze występują dwa poziomy strukturalizacji pojęć. Pierwszy z nich przyporządkowuje pojęcia do produktów PZGiK. Drugi poziom klasyfikuje pojęcia pod względem przynależności do kategorii i podkategorii tematycznych. Każde pojęcie posiada

brzmienie oraz definicję w języku polskim i angielskim, nazwę alternatywną, źródło oraz powiązane jest relacją semantyczną z daną kategorią lub podkategorią tematyczną i zbiorem danych przestrzennych PZGiK. Tezaurus został zbudowany w standardzie SKOS z wykorzystaniem przestrzeni nazw GEMET oraz Dublin Core. Zawiera 398 pojęć przyporządkowanych do poszczególnych zbiorów PZGiK oraz kategorii tematycznych. Mimo iż został stworzony na potrzeby konkretnej aplikacji, stanowi pierwszy krok w porządkowaniu terminologii dotyczącej szeroko pojętej dziedziny geoinformacji.

### **Tezaurus dla pojęć z dziedziny planowania przestrzennego**

Kolejny przykład próby ujednoczenia pojęć z danej dziedziny stanowi tezaurus zbudowany na potrzeby Wojewódzkiego Węzła Infrastruktury Informacji Przestrzennej z zakresu planowania przestrzennego. Konieczność ujednoczenia i zebrania pojęć występujących w dziedzinie szeroko pojętego planowania przestrzennego wynika nie tylko z potrzeby uporządkowania terminów, które często są różnie interpretowane. W związku z koniecznością opracowania metadanych dla danych przestrzennych, tezaurus ma służyć w opisywaniu tych zbiorów odpowiednimi słowami kluczowymi. Dzięki temu, ogranicza się dowolność w opisie metadanymi, co zwiększa możliwości wyszukiwania właściwych danych przestrzennych. Udostępniane w ramach Węzła zasoby to głównie dokumenty planistyczne, tj. studia uwarunkowań i kierunków zagospodarowania przestrzennego gmin Wrocławskiego Obszaru Metropolitalnego oraz Plan Zagospodarowania Przestrzennego Województwa Dolnośląskiego. Utworzenie tezausa, który umożliwiałby opis ww. zasobów, ze względu na różnorodność definicji jak i brak standardów w zakresie zapisu i publikacji zagadnień planistycznych, było zadaniem trudnym. Próba uporządkowania pojęć wymagała zatem zebrania i analizy wszystkich definicji i zapisów występujących w dokumentach dla analizowanego obszaru. Ostatecznie, dla potrzeb wyszukania informacji planistycznej powstał tezaurus obejmujący trzy schematy koncepcyjne, w których kolekcjonowane są pojęcia. Tezaurus został zbudowany w oparciu o standard SKOS.

## **Podsumowanie**

Porządkowanie wiedzy z danej dziedziny, z wykorzystaniem ustalonego języka do prezentacji wiedzy, stanowi etap wstępny do osiągnięcia interoperacyjności semantycznej. W aspekcie SDI istotne staje się opracowanie uzgodnionych i dzielonych modeli dla opisu zasobów geoinformacyjnych. Wynik stanowić może opracowanie wspólnego słownika, taksonomii czy też bardziej rozbudowanej struktury w postaci tezausa. Proces opracowania wspólnych pojęć i ich definicji powinien obejmować nie tylko etapy ich tworzenia, ale również ich zastosowania, kolejno utrzymywania i rozwijania. Wraz z rozwojem technologii sieci semantycznych modele organizacji wiedzy tworzy się w oparciu o standard SKOS, umożliwiając ich szersze wykorzystanie i współdzielenie.

Tezaurus w kontekście budowy infrastruktury danych przestrzennych ma za zadanie wspomagać procedury w zakresie opisu i wyszukiwania informacji przestrzennej, a także może pełnić rolę bazy zasobów terminologicznych z dziedziny geoinformacji. W związku z tym, iż celem budowy tezausa jest ułatwienie procesu indeksowania przy opisie zasobów informacyjnych, ma on istotne znaczenie w przypadku opracowania metadanych dla zbiorów danych przestrzennych. W związku, tym, że tezaury stanowią uporządkowany, za-



mniejszy zbiór, semantycznie ze sobą powiązanych pojęć, ich wykorzystanie, sprawia, że efektywność wyszukiwania w usługach katalogowych znacznie wzrasta.

### Literatura

- Rozporządzenie Komisji (WE) Nr 1205/2008 z dnia 3 grudnia 2008 r. w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie metadanych.
- IEEE Standard, 1990: Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York.
- ISO/IEC, 2003: Technical Draft Report, Ref.No. JTC1 SC36 WG4 N0070. Information Technology – Learning, education, and training – Management and delivery – Specification and use extensions and profiles. US SC36 Secretariat.
- IDABC, 2004: EIF – European Interoperability Framework for pan-European eGovernment services.
- Gaździcki J., 2002: Leksykon geomatyczny – Lexicon of Geomatics. [www.ptip.org.pl](http://www.ptip.org.pl)
- Gruber T., 1993: A translation approach to portable ontologies. Knowledge Acquisition, tom 5(2), 199-220.
- Fensel D., 2004: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag Berlin Heidelberg.
- The UK Archival Thesaurus. Dostęp 08.2011. <http://www.ukat.org.uk/>
- GEMET. Dostęp 08.2011. <http://www.eionet.europa.eu/gemet>

### Abstract

*Building spatial data infrastructure requires interoperability to be achieved at organizational, technical and semantic levels.*

*The first two topics are well recognized and implemented in practice. Problems associated with the achievement of semantic interoperability are more complex issues and still remain in the initial phase of research.*

*Semantic interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged. Semantic interoperability is crucial for further development of spatial data infrastructure and its transition to the next stage which is the spatial knowledge infrastructure.*

*The preliminary goal to achieve semantic interoperability is to organize knowledge in a given field using a formal language for presentation of knowledge. In the result, a common vocabulary, taxonomy or a more complex structure like thesaurus may be developed.*

*The traditional approach to thesaurus construction is based on the standards ISO 2788, ISO 5964 and ANSI/NISO Z39.19, describing exactly how it is created. Currently, the most widely used standard is SKOS (Simple Knowledge Organization System), developed by the W3C, consistent with the idea of ??Linked Data, which treats the Internet as a global database. In SKOS standard, thesaurus GEMET is available, which is the source of at least one of the keywords describing spatial data sets in accordance with the INSPIRE Metadata Regulation.*

*At the present stage of development of SDI, thesauri play an important role, especially in the creation of metadata, by using them as a source of keywords to describe the spatial information. Therefore, due to the fact that they form a structured closed set with semantically related concepts, the efficiency of search in catalogue services increases significantly.*

mgr inż. Iwona Kaczmarek  
[kaczmarek.iw@gmail.com](mailto:kaczmarek.iw@gmail.com)

dr inż. Adam Iwaniak  
[adam.iwaniak@gmail.com](mailto:adam.iwaniak@gmail.com)