

EDYTORY METADANYCH – PORÓWNANIE WYBRANYCH APLIKACJI „OPEN SOURCE”

METADATA EDITORS – A COMPARISON OF SELECTED OPEN SOURCE APPLICATIONS

Maciej Rossa¹, Leszek Litwin²

¹ Państwowy Instytut Geologiczny

² Instytut Systemów Przestrzennych i Katastralnych S.A.

Słowa kluczowe: metadane, edytor metadanych, standard otwarty, INSPIRE
Keywords: metadata, metadata editor, open standard, INSPIRE

Wstęp

Metadane, czyli dane o danych, są kluczowym, podstawowym i niezbędnym elementem funkcjonowania infrastruktur danych przestrzennych. Przede wszystkim ułatwiają one zarządzanie zgromadzonymi zasobami i korzystanie z nich oraz umożliwiają wyszukiwanie konkretnych i interesujących nas danych i usług (Gaździcki, 2003; Iwaniak, 2005). Ponadto opracowanie oraz stosowanie metadanych jest obowiązkowe w krajach Unii Europejskiej. Obowiązek ten nałożyła dyrektywa INSPIRE z dniem 15.05.2007 r. Określa ona m.in. terminy, do których należy przygotować odpowiednie metadane: do 15.05.2009 r. dla zasobów wymienionych w załącznikach I i II dyrektywy oraz do 15.05.2013 r. dla zasobów z załącznika III. Ponieważ dyrektywa INSPIRE jest dokumentem o charakterze ogólnym, dlatego do praktycznego jej stosowania (w tym tworzenia metadanych) niezbędne są szczegółowe wytyczne.

Wytyczne takie na poziomie europejskim to Zasady Wdrażania (*IR – Implementing Rules*). Dla metadanych zostały one przygotowane przez Grupę Roboczą ds. metadanych (*DT – Drafting Team*) i przyjęte przez Komisję Europejską w dniu 14.05.2008 r. Ostateczną formą tych wytycznych jest Rozporządzenie Komisji Wspólnot Europejskich nr 1205/2008 z dnia 03.12.2008 r., ustanawiające wymagania w zakresie tworzenia i przechowywania metadanych dla zbiorów danych przestrzennych, serii zbiorów danych przestrzennych i usług danych przestrzennych dotyczących tematów wymienionych w załącznikach I, II i III do dyrektywy. Jest to dokument bezpośrednio wiążący i obowiązujący wszystkie kraje członkowskie. Wchodzi ono w życie dwudziestego dnia po jego opublikowaniu w Dzienniku Urzędowym UE.

Równoległe z przyjęciem wytycznych INSPIRE dla metadanych, na stronie internetowej Głównego Urzędu Geodezji i Kartografii (GUGiK) (<http://www.gugik.gov.pl>) został opubli-

kowany (w dniu 07.05.2008 r.) pod linkiem „Polski krajowy profil metadanych w zakresie geoinformacji” pakiet dokumentów obejmujący: opis krajowego profilu metadanych, opis architektury systemu metadanych oraz odpowiednie wytyczne techniczne. Należy zaznaczyć, że dokumenty te są w większości zgodne z wytycznymi INSPIRE, ale obecnie mają tylko status standardu otwartego. Definicja standardu otwartego dostępna jest w Wikipedii http://pl.wikipedia.org/wiki/Otwarty_standard. Wynika z niej, że dokumenty te nie stanowią formalnych wytycznych krajowych, a ich stosowanie jest dobrowolne.

Podstawą opracowania wytycznych europejskich i krajowych, są światowe standardy geoinformacyjne, przede wszystkim normy: ISO 19115:2003/Cor1:2006 *Metadata* oraz ISO/TS 19139:2007 *Metadata – XML schema implementation*. Norma ISO 19115 jest także obowiązującą normą europejską CEN i polską PN.

Wymienione powyżej wytyczne i normy pozwalają w pełni na rozpoczęcie procesu tworzenia metadanych dla INSPIRE w naszym kraju. Aby proces ten przebiegał sprawnie, a jego postęp był znaczący, należy do edycji metadanych używać odpowiednich narzędzi informatycznych, przede wszystkim edytorów metadanych.

Głównym celem niniejszej pracy była analiza porównawcza wybranych edytorów metadanych pod względem możliwości ich zastosowania w procesie edycji metadanych dla polskiej infrastruktury danych przestrzennych zgodnie z wytycznymi krajowymi i INSPIRE. Ponieważ Komisja Europejska zaleca stosowanie i rozwijanie oprogramowania typu *Open Source* (o otwartym kodzie i przeważnie darmowe), w pracy przedstawiono i porównano cztery edytory metadanych należące do tej grupy: CatMDEdit, I.M.E., MEDARD i MEE.

Prezentowane w pracy wyniki są efektem działań podejmowanych przez Państwowy Instytut Geologiczny (PIG) na rzecz wdrażania INSPIRE przez państwowe służby geologiczną i hydrogeologiczną, w tym współpracy w ramach Sieci Naukowej „Systemy Geoinformacyjne” z 12 głównymi jednostkami badawczo-rozwojowymi w kraju oraz Instytutem Systemów Przestrzennych i Katastralnych S.A. z Gliwic (ISPiK). Współpraca ta obejmuje głównie analizy możliwości implementacji różnych profili metadanych (w tym geologicznego PIG oraz dedykowanego dla Sieci), edycji metadanych według profili społecznościowych (geologiczny, hydrogeologiczny, hydrologiczny itp.) oraz budowy i rozwoju polskiego standardowego edytora metadanych MEDARD.

Wymagania stawiane edytorom metadanych

W celu przeprowadzenia analizy porównawczej wybranych edytorów metadanych, w pierwszej kolejności zdefiniowano pojęcie edytor metadanych, jako narzędzie informatyczne pozwalające twórcom metadanych na przygotowanie zbioru metadanych w określonym, przyjętym standardzie. Następnie przeanalizowano przytoczone powyżej dokumenty z punktu widzenia postawionego zadania. Na tej podstawie sporządzono poniższe zestawienie wymagań stawianych edytorom metadanych, według których powinny one:

- zapewniać pełną obsługę języka polskiego;
- obsługiwać metadane zgodnie z profilami: ISO 19115, INSPIRE i polskim (krajowym), ewentualnie branżowym (np. profil metadanych geologicznych PIG);
- umożliwiać tworzenie i aktualizację metadanych w określonej hierarchii z możliwością dziedziczenia elementów metadanych;

- automatycznie generować identyfikatory plików metadanych, zgodnie z UUID (*Universal Unique Identifier*, specyfikowany przez IETF <http://www.ietf.org> oraz RFC 4122);
- umożliwiać dostosowanie funkcjonalności oprogramowania do wymagań zawartych w wytycznych technicznych danego profilu;
- zapewniać możliwość pobierania słowników słów kluczowych i kategorii tematycznych z odpowiednich baz danych (*off-line* i *on-line*);
- umożliwiać budowę wersji wielojęzycznych metadanych przy wykorzystaniu kodowania z ISO 639-2;
- zapewniać możliwość wprowadzanie informacji dla zbiorów oraz serii danych;
- generować pliki metadanych w formacie XML zgodnie ze schematem implementacyjnym (XML Schema) określonym w standardzie ISO/TS 19139:2007;
- zapisywać elementy metadanych opisujące zasięgi czasowe zgodnie ze standardem GML, opisanym w dokumencie ISO/TS 19139:2007. Konwersja do tego standardu powinna odbywać się automatycznie przy wykorzystaniu mechanizmów edytora metadanych;
- ułatwiać definiowanie zasięgów przestrzennych zasobu, przy czym mogą one być zdefiniowane na kilka sposobów, jako:
 - prostokąt ograniczający zapisany w postaci jego boków (skrajnych),
 - opis w postaci swobodnego tekstu,
 - poligon zapisany w formie współrzędnych (x, y) wieloboku ograniczającego opisywany zasób – zapisany zgodnie ze standardem GML opisanym w dokumencie ISO/TS 19139:2007. Konwersja do tego standardu, powinna odbywać się automatycznie przy wykorzystaniu mechanizmów edytora metadanych,
 - identyfikator geograficzny w formie odniesienia (identyfikatora) do obiektów/obiektu geograficznych posiadających lokalizację przestrzenną i opublikowanych w udokumentowanej i ogólnie dostępnej usłudze gazetteer’a lub katalogu obiektów przestrzennych.
- umożliwiać automatyczne generowanie odniesienia do sekcji informacji o identyfikacji odpowiedniego pliku metadanych, który opisuje zasób danych sprzężony z usługą.

W tym miejscu należy dodać, że edytory metadanych służące dla opisu metadanych zasobów należących do szeroko rozumianych Nauk o Ziemi (np. geologia, górnictwo, archeologia, gleboznawstwo, geotechnika itp.) powinny również umożliwiać edycję wybranych elementów metadanych zgodnie ze standardem języka GeoSciML (*Markup Language for the Geosciences*), który jest aplikacją języka GML. Przykładem może być tu, wymagany w metadanych INSPIRE, opis wieku geologicznego zgodnie z tabelą stratygraficzną, jako przestrzeni czasu, której dotyczą opisywane zasoby. GeoSciML jest obecnie standardem OGC (*Open Geospatial Consortium*), ale z dużym prawdopodobieństwem można dziś stwierdzić, że stanie się on obowiązującym standardem INSPIRE w zakresie geologii i hydrogeologii, zagrożeń naturalnych, zasobów mineralnych i energetycznych oraz gleb.

Opis analizowanych edytorów metadanych

Analizie porównawczej poddano cztery edytory metadanych z grupy oprogramowania OpenSource. Edytory CadMDEdit i I.M.E. to wyniki projektów europejskich, natomiast edytory MEDARD i MEE to oprogramowanie opracowane w Polsce. W celu ogólnej prezentacji i opisu analizowanych edytorów sporządzono zestawienie ich głównych cech i parametrów, które przedstawiono w formie tabelarycznej (tab. 1).

Wyniki analizy porównawczej edytorów metadanych

Na wstępie należy podkreślić, że wszystkie cztery omawiane edytory metadanych odpowiadają przyjętej definicji. Natomiast nie wszystkie spełniają wymagania stawiane edytorom metadanych, które przedstawiono powyżej.

Jednoczesną obsługę profili: ISO, INSPIRE oraz krajowego zapewnia tylko jeden edytor – MEDARD. Ponadto, tylko MEDARD zapewnia możliwość tworzenia i aktualizacji metadanych w określonej hierarchii z możliwością dziedziczenia elementów metadanych. W edytorach CatMDEdit oraz MEE można budować hierarchie, ale bez obsługi dziedziczenia. Oprócz edytora I.M.E., pozostałe edytory automatycznie generują identyfikatory zgodnie z UUID. Tylko najbardziej rozbudowane i zaawansowane edytory, takie jak CatMDEdit i MEDARD, zapewniają możliwość pobierania słowników słów kluczowych i kategorii tematycznych z odpowiednich baz danych. Edytor MEE też posiada możliwość podłączenia słowników, ale są to wewnętrzne słowniki aplikacji zapisane w języku XML (są one edytowalne). Za pomocą wszystkich omawianych edytorów można tworzyć metadane w różnych językach, ale tylko MEDARD posiada narzędzia do budowy wielojęzycznych metadanych przy wykorzystaniu kodowania z ISO 639-2. Wszystkie prezentowane edytory zapewniają możliwość wprowadzania informacji zarówno dla zbiorów danych, serii danych oraz serwisów. Również każdy z nich może generować pliki metadanych w formacie XML zgodnie ze schematem implementacyjnym (XML Schema) określonym w standardzie ISO/TS 19139:2007, jednak należy zwrócić szczególną uwagę według której z wersji pliku XSD następuje walidacja. Serwis MEDARD'a oficjalnie zapewnia najnowszą obowiązującą wersję tego pliku, MEE ma możliwość podania lokalizacji odpowiedniej wersji tego pliku. Pozostałe dwa edytory korzystają z plików XSD, które są „na sztywno” zapisane w kodzie programu – ich aktualizacja wymaga ingerencji w kod oprogramowania. Tylko edytor MEDARD zapewnia zapis elementów metadanych opisujące zasięgi czasowe zgodnie ze standardem GML, opisanym w dokumencie ISO/TS 19139:2007. Ponadto umożliwia opis tych elementów w aplikacji języka GML – GeoSciML. Oprócz edytora I.M.E., pozostałe edytory ułatwiają definiowanie zasięgów przestrzennych zasobu, w formie: prostokąta ograniczającego, opisu tekstowego, poligonu zapisanego w formie współrzędnych (x, y), czy identyfikatora geograficznego z usługi gazetteer'a. Najbardziej zaawansowanym w tym względzie jest CatNDEdit, który oferuje nawet możliwość podglądu zasięgu zasobu na warstwach przestrzennych (mapach). Każdy z edytorów posiada mniej lub bardziej obszerną i szczegółową instrukcję użytkownika oraz pomoc, jednak najbardziej przyjaznym wydaje się być MEDARD, który posiada, zarówno odpowiednią dokumentację, pomoc kontekstową w aplikacji, jak wideoprezentacje czy FAQ.

Dla pełniejszego i bardziej przejrzystego zaprezentowania wyników przeprowadzonej analizy porównawczej zestawiono je w formie tabelarycznej (tab. 2).

Tabela 1. Zestawienie głównych cech i parametrów czterech edytorów

Cecha/ parametr	CatMDEdit	I.M.E.	MEDARD	MEE
Strona internetowa projektu	http://catmdedit.sourceforge.net/	http://www.crepad.rcanaria.es/metadata/index.htm	http://medard-opensource.eu/	https://sourceforge.net/projects/metaddata/monitor/
Aktualna wersja	4.0.1 (08.04.2008)	4.1 (08.2007) – MS Windows 4.0 (09.2006) – Linux	1.1.1 (11.2008)	1.04 RC1 (02.2008) – RC (Release Candidate)
Typ licencji	GNU 2.1	OpenSource	GNU AGPL v3	LGNU
Twórcy	University of Zaragoza GeoSpatialLab S.L.	INTA (National Institute for Aerospace Technology)	ISPIK (Instytut Systemów Przeszrzynnych i Katastralnych S.A. z Gliwic)	Bartosz Kopaczek na zamówienie GUGiK
System operacyjny	wirtualna maszyna Java	MS Windows Linux	wirtualna maszyna Java	MS Windows (32 bit) Linux
Obsługiwane języki	angielski, czeski, hiszpański, francuski, polski, portugalski	angielski, hiszpański, polski	angielski, polski	angielski, polski
Zarządzanie metadany	poprzez interfejs z lokalnego repozytorium	brak	poprzez interfejs użytkownika lub poprzez WWW z repozytorium lokalnego lub sieciowego oraz z bazy danych; możliwość pracy wielostanowiskowej	brak
Obsługiwane bazy danych	lokalne repozytorium plików	brak	dowolne RDBMS	brak informacji
Zgodność z geostandardami	ISO 19115:2003/Cor1:2006 ISO 19139:2007	ISO 19115:2003/Cor1:2006 ISO 19139:2007	ISO 19115:2003/Cor1:2006 ISO 19136:2004 ISO 19139:2007 OGC GeoSciML 2 i 3 ISO 19103	ISO 19115:2003/Cor1:2006 ISO 19139:2007
Obsługiwane profile	ISO 19115 INSPIRE Dublin Core WISE (Ramowa Dyrektywa Wodna) NEM (Núcleo Español de Metadatos)	ISO 19115 (miejscami ograniczony)	ISO 19115 INSPIRE polski (krajowy GUGiK) PIG (geologiczny)	polski (krajowy GUGiK)
Możliwość tworzenia własnych profili metadanych	istnieje, ale brak narzędzi w menu	tak	tak	brak

cd. tabeli 1

Cecha/ parametr	CatMDEdit	I.M.E	MEDARD	MEE
Dziedziczenie	brak	brak	zawiera	brak
Możliwość tworzenia szablonów (wzorców)	brak	tak	tak	brak
Narzędzia do walidacji	zawiera	zawiera	zawiera	zawiera
Podgląd metadanych	HTML	XML, HTML (bezpośrednio – automatyczne generowanie plików lub pośrednio – tworzenie pliku transformacji XSLT) PDF (raporty)	XML, HTML, PDF	brak
Import/ eksport	XML (ISO19139, CSD GM-US FGDC, SDIGER) i RDF (Dublin Core)	XML (ISO19139)	XML (ISO19139)	XML (ISO19139)
Automatyczne generowanie metadanych	Shapefile, DGN, ECW, FICC, GeoTiff, GIF/GFW, JPG/JGW, PNG/PGW	brak	brak	brak
Dodatkowe narzędzia ułatwiające edycję metadanych	<ul style="list-style-type: none"> – Repozytorium kontaktów, wielokrotne wykorzystanie danych teleadresowych. – Thesaurus, wypełnienie elementów metadanych z podłączonych słowników. – Konwersja pomiędzy różnymi systemami współrzędnych. – Narzędzia do edycji, kontroli i wizualizacji (możliwy podgląd na mapie) elementów opisujących zakres przestrzenny zasobów w postaci poligonów. – Wizualizacja danych i miniatur w różnych formatach: Shapefile, ECW, GeoTiff, GIF, JPG, BMP, PDF, HTML. 	Wskazanie powiązań elementów metadanych do innych standardów: ISO19103, ISO10107, ISO19108, ISO19109, ISO19136 (GML)	<ul style="list-style-type: none"> – Repozytorium kontaktów, wielokrotne wykorzystanie danych teleadresowych (import formatu CSV). – Thesaurus, wypełnienie elementów metadanych z podłączonych słowników. – Narzędzia do wyszukiwania pełnotekstowego. – Pomoc kontekstowa (edycja, obsługa), instrukcje, prezentacje, video prezentacje oraz FAQ. – Narzędzia do edycji, kontroli i wizualizacji (możliwy podgląd) elementów opisujących zakres przestrzenny zasobów w postaci poligonów. 	<ul style="list-style-type: none"> – Narzędzia do edycji elementów opisujących zakres przestrzenny zasobów w postaci poligonów. – Narzędzia do edycji elementów opisujących typ prezentacji przestrzennej.

Tabela 2

Wymagania	CatMDEdit	I.M.E.	MEDARD	MEE
Pełna obsługa języka polskiego	+	+	+	+
Obsługa profili ISO, INSPIRE, krajowego	–	–	+	–
Obsługa hierarchii metadanych z dziedziczeniem	–	–	+	–
Automatyczne generowanie identyfikatorów metadanych wg UUID	+	–	+	+
Pobieranie słowników słów kluczowych i kategorii tematycznych	+	–	+	+(tylko wewnętrzne słowniki)
Tworzenie wielojęzycznych wersji metadanych	–	–	+	–
Obsługa metadanych dla serii i zbiorów	+	+	+	+
Generowanie plików XML zgodnie z ISO/TS 19139:2007	+	+	+(aktualne XSD)	+(aktualne XSD)
Opis czasu zgodnie z GML	–	–	+	–
Definiowanie zasięgów przestrzennych jako:		Możliwe ale brak ułatwień edycyjnych		
– prostokąt ograniczający	+	+	+	+
– tekst	+	+	+	+
– poligon (GML)	+	+	+	+
– identyfikator	+	+	+	+
Implementacja GeoSciML	–	–	+	–

Podsumowanie

Jak wynika z przeprowadzonej analizy porównawczej omawianych edytorów, tylko edytor MEDARD spełnia wszystkie postawione wymagania. Ponadto, jako jedyny obsługuje dziedziczenie oraz implementuje języki GML i GeoSciML. Posiada także bardzo rozbudowaną stronę wsparcia dla użytkownika (instrukcje, prezentacje, video prezentacje, FAQ). Jest to również jedyny edytor metadanych w tej grupie, którego projekt jest nadal czynnie kontynuowany, wspierany i rozwijany. W tym miejscu należy podkreślić, że cechy i parametry edytora MEDARD zostały dostrzeżone przez Państwowy Instytut Geologiczny (gdzie jest wykorzystywany do edycji metadanych przez państwową służbę geologiczną i hydrogeologiczną) oraz partnerów Sieci Naukowej „Systemy Geoinformacyjne”. Ponadto, wykorzystaniem edytora MEDARD zainteresowane są również jednostki związane z Ministerstwem Środowiska, takie jak Główny Inspektorat Ochrony Środowiska, Natura 2000, Instytut Meteorologii i Gospodarki Wodnej czy Krajowy Zarząd Gospodarki Wodnej. Edytor MEDARD był również prezentowany na forum projektu UE OneGeologyEurope jako pierwszy na świecie edytor metadanych implementujący język GeoSciML, gdzie został przyjęty z dużym zainteresowaniem.

Oprócz MEDARD’a, edytorem implementującym profil krajowy jest MEE – jednakże edytor ten nie realizuje dziedziczenia, co jest wymagane w wytycznych. Ponadto, projekt

edytora MEE jest obecnie zamknięty i nie jest czynnie wspierany, a możliwości jego kontynuacji są dziś niewiadomą.

Edytor CatMDEdit jest produktem bardzo zaawansowanym i dojrzałym, ale nie implementuje polskiego profilu krajowego. Dodanie tego profilu jest możliwe, ale wymaga ingerencji w kod oprogramowania i nie jest możliwe do przeprowadzenia przez zwykłego użytkownika CatMDEdit. Należy jednak zaznaczyć, że edytor ten posiada zaimplementowanych wiele różnych profili metadanych związanych z innymi przepisami UE. Ponadto ma najbardziej rozbudowaną funkcjonalność i zaawansowany interfejs użytkownika, ułatwiający definiowanie zasięgów przestrzennych zasobu.

Najmniej zaawansowanym edytorem i jednocześnie spełniającym najmniej wymogów jest edytor I.M.E. Zasluguje on jednak na uwagę gdyż bardzo ułatwia poznanie, zrozumienie i pracę ze standardami ISO19115 i ISO19139.

Literatura

- Gaździcki J., 2003: Kompendium infrastruktur danych przestrzennych, cz. II, Metadane. *Magazyn Geoinformacyjny Geodeta* 3.
- Iwaniak A., 2005: Infrastruktura danych przestrzennych (SDI) inaczej, cz. I. *Magazyn Geoinformacyjny Geodeta* 11.
- Implementing Rule for Metadata. <http://inspire.jrc.it>
- ISO 19115:2003/Cor1:2006 *Metadata*
- ISO/TS 19139:2007 *Metadata – XML schema implementation*
- ISO 19139:2007 *Metadata – XML schema implementation*
- Polski krajowy profil metadanych w zakresie geoinformacji. <http://www.gugik.gov.pl>
- Strona internetowa projektu CatMDEdit <http://catmdedit.sourceforge.net/>
- Strona internetowa projektu I.M.E. <http://www.crepad.rcanaria.es/metadata/index.htm>
- Strona internetowa projektu MEDARD <http://medard-opensource.eu/>
- Strona internetowa projektu MEE <http://sourceforge.net/projects/metadataeditor/>

Abstract

In the introduction we discuss the need and obligations concerning elaboration and use of metadata resulting from the INSPIRE Directive and secondary provisions. We presented also ISO standards and Polish guidelines and standards, which allow beginning the process of creating metadata in Poland. For its efficient course, appropriate metadata editors are necessary.

We present in the paper: 1) requirements posed for metadata editors; 2) description of analyzed 4 selected editors of the Open Source type 4 presented in Table 1 (elaborated within the framework of European programmers – CatMDEdit and I.M.E. and elaborated in Poland – MEDARD and MEE); 3) results of comparative analysis presented in Table 2.

The metadata editors were analyzed with respect to the possibility of using them in the process of metadata editing for Polish spatial information infrastructure in accordance with national and INSPIRE guidelines. The editors result from actions undertaken by the Polish Geological Institute (PIG) for implementation of INSPIRE by state geological and hydrogeological services, including cooperation within the framework of Scientific Network „Geoinformation Systems” with a number of research and development institutes in the country. This cooperation comprises mostly the analyses of the possibilities to implement various metadata profiles (including geological one), metadata editing according to community profiles (geological, hydrogeological, hydrological etc.) and development of Polish metadata editor MEDARD.

mgr Maciej Rossa
maciej.rossa@pgi.gov.pl