

METADANE W REGIONALNYM SYSTEMIE INFORMACJI PRZESTRZENNEJ

METADATA IN A REGIONAL SPATIAL INFORMATION SYSTEM

Edward Liszczyk

FORMIDA Laboratorium Informatyki

Słowa kluczowe: system informacji przestrzennej, regionalny system informacji przestrzennej, metadane, architektura, ISO

Keywords: Spatial information system, regional spatial information system, metadata, architecture, ISO

Streszczenie

Prace badawcze dotyczące metadanych zostały podjęte w latach osiemdziesiątych ubiegłego wieku. Były one spowodowane lawinowym wzrostem nagromadzonych zasobów danychi przestrzennych. Metadane przyczyniają się do uporządkowania tych zasobów, stanowiąc ich sformalizowaną dokumentację. Intensywne prace w wielu krajach doprowadziły do powstania interesujących rozwiązań implementacyjnych, które przyczyniły się do powstania międzynarodowej normy ISO 19115. W Polsce prace nad tym zagadnieniem podjęto m.in. w Instytucie Systemów Przestrzennych i Katastralnych S.A w związku z projektami dotyczącymi regionalnego systemu informacji przestrzennej (RSIP). Opracowano architekturę podsystemu metadanych RSIP z użyciem nowoczesnych technik projektowania i w dostosowaniu do architektury całego systemu. Powstały podsystem metadanych jest elastyczny i zgodny z normą ISO 19115, także pod względem budowania specyficznych profili i rozszerzeń. Wychodzi on również naprzeciw zaleceniom INSPIRE.

Wstęp

Jak wiadomo metadane są danymi o danych. Analogicznie do dokumentacji różnych produktów rynkowych metadane stanowią dokumentację zbiorów danych. W tym ujęciu zbiór danych jest traktowany jako produkt, a metadane stanowią jego dokumentację.

Metadane nabierają znaczenia wraz ze wzrostem liczby danych, jakie wymagają opisu. Proste i niewielkie zasoby danych mogą funkcjonować bez dokumentacji w formie metadanych. W wypadku dużych zasobów i złożonych powiązań między danymi metadane są praktycznie niezbędne.

Metadane wspierają użytkownika w trzech obszarach działania:

- wyszukiwanie informacji – podsystem metadanych zapewnia możliwość łatwego przeszukiwania metadanych na podstawie atrybutów przestrzennych i nieprzestrzennych określających poszukiwane dane,
- pobieranie danych – za pomocą podsystemu metadanych potencjalny użytkownik może zostać skierowany do poszukiwanych zasobów, aby uzyskać dostęp do danych,
- wykorzystanie danych – metadane pozwalają użytkownikowi upewnić się, czy znalezione dane są odpowiednie i wystarczające z punktu widzenia użytkownika..

Podsystem metadanych udostępnia dokumentację zasobów systemu informacji przestrzennej na różnych poziomach agregacji i abstrakcji. Każdy rekord metadanych opisujący dowolny zasób powinien dawać użytkownikowi odpowiedzi na sześć podstawowych pytań o dane:

- **Co?** – nazwa zbioru danych, opis danych
- **Kto?** – nazwa i inne dane kontaktowe dysponenta danych odpowiedzialnego za ich utrzymanie i udostępnianie,
- **Gdzie?** – określenie przestrzennego położenia obiektów, lub obszarów których dotyczą dane. Położenie może być określone współrzędnymi, nazwami obszarów geograficznych i administracyjnych lub też przez opis słowny,
- **Po co?** – szczegóły dotyczące przeznaczenia danych oraz ich potencjalnego wykorzystania,
- **Kiedy?** – kluczowe daty dotyczące metadanych – data założenia rekordu, data modyfikacji, planowane zmiany, etc.,
- **Jak?** – sposoby pobierania danych z opisywanego zasobu i zasady dostępu do nich. Metody dostępu do danych obejmują również dostęp do danych nie mających reprezentacji numerycznej.

Ze względów praktycznych dogodnie jest ujęcie metadanych w grupy tematyczne, co w sposób przykładowy pokazano poniżej.

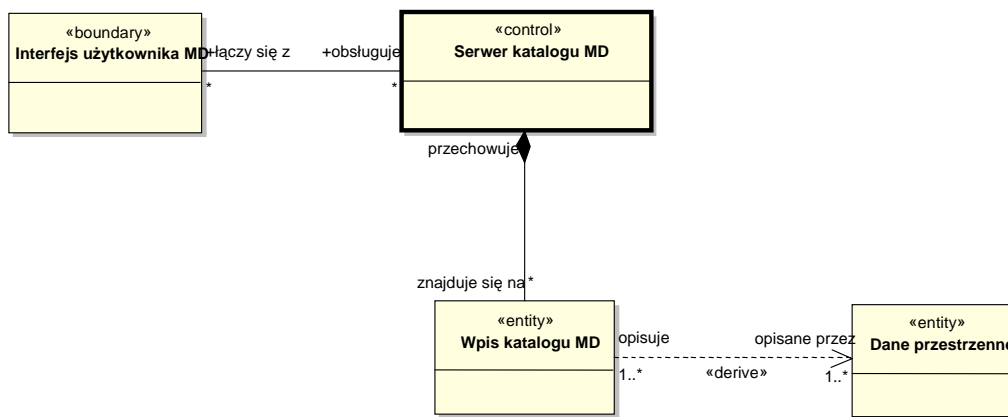
- Metametadane – opis metadanych zawierający ich identyfikację, kontakt, informacje o wykorzystanym standardzie.
- Identyfikacja – nazwa i opis zbioru danych, słowa kluczowe, dane dysponenta opisywanych danych, pokrywany obszar geograficzny.
- Jakość – dokładność określenia współrzędnych, kompletność danych, informacja czy spójność danych była weryfikowana, na podstawie jakich danych zbior został utworzony i jaki proces został zastosowany.
- Organizacja przestrzenna – model przestrzenny, któremu dane odpowiadają.
- Opis danych przestrzennych – tematyka, obiekty, atrybuty oraz ich dziedziny, kodowanie.
- Informacja dystrybucyjna – gdzie można uzyskać dane, w jakich formatach, za jaką cenę.
- Informacja o osobach i instytucjach odpowiedzialnych za metadane i dane.

W artykule niniejszym przedstawia się wyniki prac wykonanych w Polsce w ramach projektów prowadzonych przez Instytut Systemów Przestrzennych i Katastralnych S.A. i dotyczących metadanych w regionalnych systemach informacji przestrzennej (RSIP).

Architektury podsystemów metadanych w regionalnych systemach informacji przestrzennej

W trakcie projektowania podsystemu metadanych na potrzeby regionalnych systemów informacji przestrzennej rozważano różne modele architektury, poszukując rozwiązania optymalnego. Brano pod uwagę następujące typowe rozwiązania:

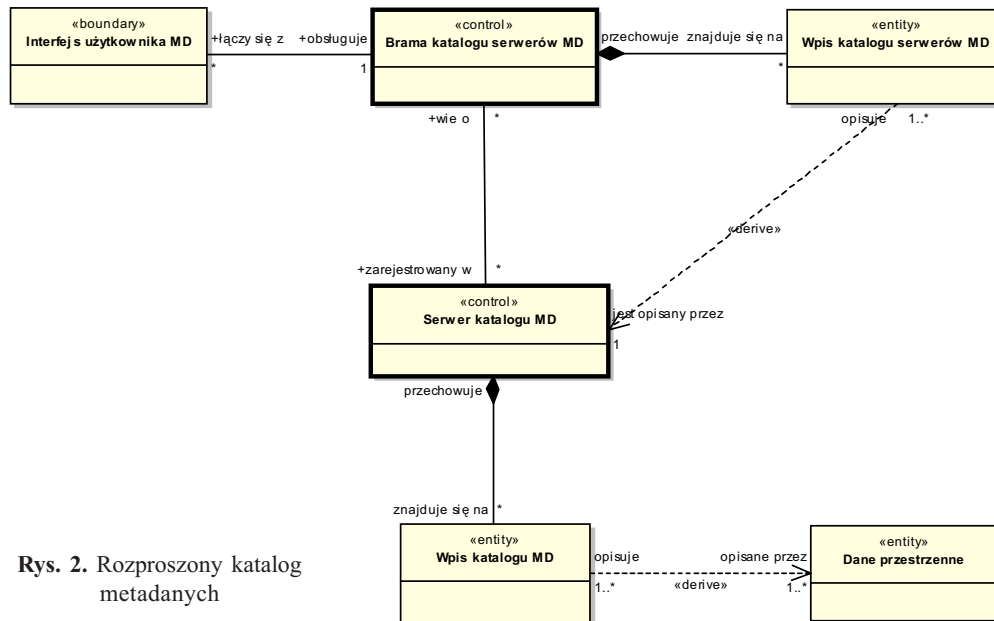
Niezależny serwer katalogu metadanych. Diagram klas tego modelu prezentuje rysunek 1. W modelu tym użytkownik przez interfejs użytkownika łączy się z wybranym serwerem metadanych. W takiej architekturze występuje wiele niezależnych serwerów metadanych co rodzi konieczność łączenia się z każdym z nich osobno. Nie daje to możliwości łączenia metadanych pochodzących z różnych źródeł.



Rys. 1. Niezależny serwer katalogu metadanych

Rozproszony katalog metadanych. Model tej architektury przedstawia rysunek 2. Takie podejście ma na celu przezwycięzenie problemów występujących w modelu z niezależnymi serwerami katalogów metadanych, przy jednoczesnym zachowaniu ich autonomii. Użytkownik poprzez interfejs łączy się z bramą rozproszonego katalogu metadanych i przesyła do niej zapytania. Brama na podstawie metadanych opisujących serwery określa, do których serwerów katalogu metadanych należy przekazać zapytanie. Jest ono rozsyłane do wybranych serwerów i wykonywane na nich lokalnie. Otrzymane odpowiedzi są scalane i odsyłane do użytkownika.

Brama umożliwia rejestrowanie nowych serwerów metadanych, aktualizację opisów zarejestrowanych oraz usuwanie serwerów z rejestru. Aby całość mogła funkcjonować, wszystkie serwery katalogu metadanych muszą rozumieć przesyłane do nich zapytania i odsyłać odpowiedzi w formacie zrozumiałym dla bramy. W praktyce wykorzystuje się standard wymiany informacji Z39.50. Przedstawiona koncepcja jest szczególnie przydatna, jeżeli metadane są udostępniane przez różne równoprawne organizacje, które chcą udostępniać dane w oparciu się o ten sam profil.



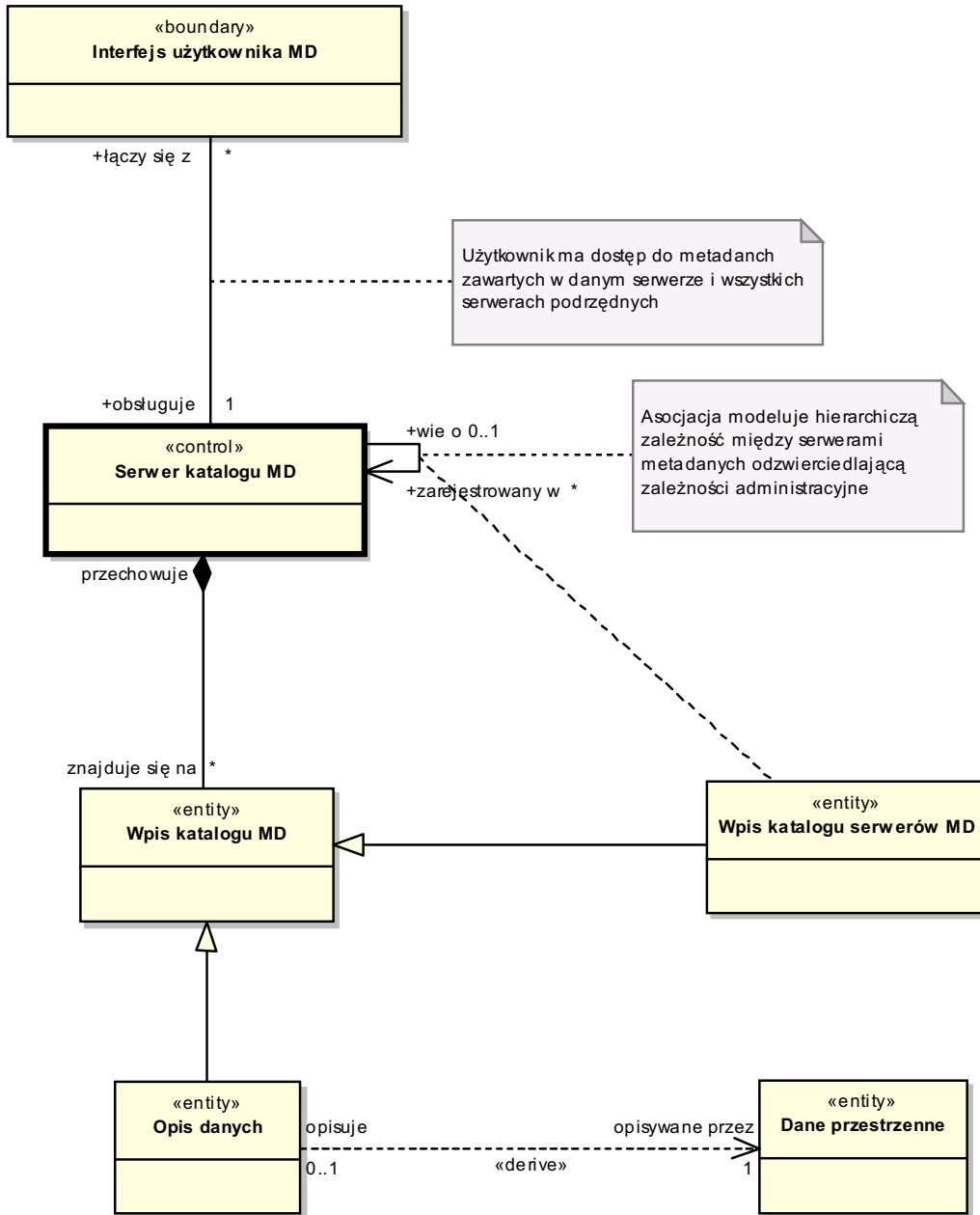
Rys. 2. Rozproszony katalog metadanych

Rozproszony katalog metadanych z wbudowaną bramą. Model tego rozwiązania jest przedstawiony na rysunku 3. Rozproszony katalog metadanych z wbudowaną bramą stanowi modyfikację modelu rozproszonego katalogu metadanych opisanego wyżej. W rozwiązaniu tym serwery metadanych są połączone hierarchicznie i każdy z nich stanowi bramę do poddrzewa, którego jest korzeniem. Wyróżniono dwa podtypy wpisów w katalogu metadanych: opisujące dane i opisujące serwery. Tworzą one dwa logicznie osobne katalogi metadanych. Metadane opisujące serwery określają hierarchiczne powiązanie serwerów oraz mogą determinować, do których podrzędnych serwerów należy odesłać dane zapytanie. Model ten pozwala na ukrycie fizycznej organizacji metadanych na poziomie instytucji – niezależnie od tego, czy metadane są przechowywane na jednym serwerze, czy na wielu, na wyższym poziomie widać już tylko jeden serwer metadanych, który pozwala jednak na dostęp do wszystkich metadanych na niższych poziomach. W rozwiązaniu tym można utworzyć usługę, która polega na rekurencyjnym przejrzaniu wpisów w katalogu serwerów i utworzeniu na tej podstawie listy serwerów.

Podsystem metadanych w RSIP

Wybierając model architektury dla podsystemu przeglądania i przeszukiwania metadanych dla praktycznej realizacji w regionalnym systemie dążono do:

- spełnienia wymagań funkcjonalnych dotyczących przeglądania i przeszukiwania oraz edycji metadanych,
- umożliwienia pracy w rozproszonym środowisku heterogenicznym,
- powiązania danych z ich dokumentacją, czyli metadanymi,



Rys. 3. Rozproszony katalog metadanych z wbudowaną bramą

- stworzenia możliwości przyszłej integracji podsystemu z infrastrukturą danych przestrzennych na poziomie krajowym i międzynarodowym,
- zgodności z architekturą wybraną dla całego systemu informacji przestrzennej.

Zdecydowano się na wybór architektury z rozproszonym katalogiem metadanych i wbudowaną bramą, ponieważ lepiej modeluje ona zależności hierarchiczne występujące między jednostkami administracji. Umożliwia dołączenie jednostek branżowych z jednoczesnym ukryciem wewnętrznej organizacji ich serwerów. Niezależnie od ilości serwerów metadanych w ośrodku, jest on widziany na zewnątrz jako jeden katalog, który można przeglądać i przeszukiwać. Tworzenie hierarchicznej struktury serwerów w przeciwieństwie do płaskiej listy pozwala w naturalny sposób ograniczać warunki zapytania oraz zmniejszać obciążenie sieci i serwerów. Nie ogranicza ono użytkownika, który zawsze może przełączyć się do nadrzędnego serwera, aby poszerzyć zakres przeszukiwanych serwerów. Wadą tego rozwiązania jest odcięcie całego poddrzewa w przypadku awarii węzła wewnętrznego lub linii komunikacyjnej. Jednakże poddrzewa mogą dalej pracować.

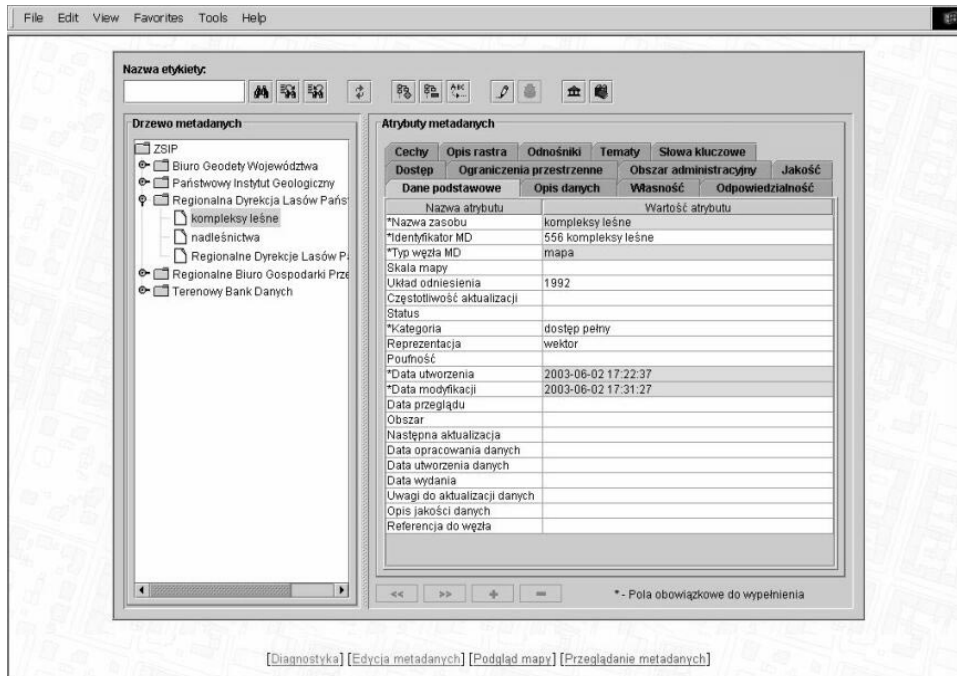
Podsystem metadanych jest składową całego regionalnego systemu informacji przestrzennej. Został on zatem dostosowany do architektury tego systemu. W RSIP zastosowano architekturę wielowarstwową. Głównym elementem tej architektury jest serwer danych przestrzennych ISDP. Stanowi on podstawę architektoniczną wszystkich aplikacji składowych RSIP, w tym również aplikacji dla podsystemu metadanych.

Funkcjonalność podsystemu metadanych zapewniają dwie aplikacje: edytora metadanych i dostępu do metadanych. Obie aplikacje zostały zbudowane jako rozszerzenia ISDP zgodnie z wymaganiami tego modułu. Jako interfejs użytkownika (GUI) został zastosowany aplet uruchamiany w środowisku przeglądarki internetowej. Wybór tej techniki został podyktowany faktem, iż zadania stojące przed aplikacją użytkownika przekraczają możliwości prostych technik (HTML, JavaScript). Warto dodać, że mimo złożoności samego GUI dotyczącego prezentacji i edycji metadanych podsystem posiada wszelkie cechy architektury *thin client*. Całe przetwarzanie danych odbywa się bowiem po stronie serwera.

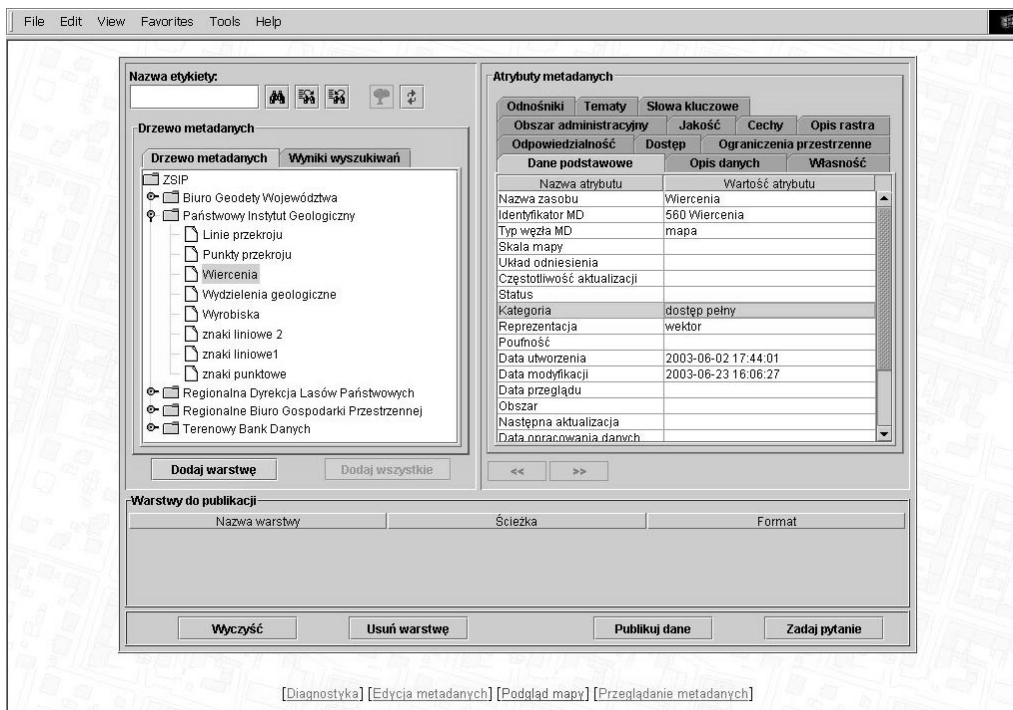
Aplikacje podsystemu metadanych

Aplikacja edytora metadanych. Edytor metadanych jest zarazem ich uproszczoną przeglądarką. Trudno sobie bowiem wyobrazić edycję metadanych bez możliwości ich jednoczesnej obserwacji. Edytor metadanych zapewnia użytkownikowi funkcje tworzenia, przesuwania, usuwania i modyfikacji rekordów metadanych. Aplikacja użytkownika przetwarza polecenia użytkownika, kolekcjonuje dane i komunikuje się z serwerem, który dokonuje odpowiednich operacji na bazie danych. Serwer również dba o pełną synchronizację danych i nie dopuszcza do niekontrolowanych, jednoczesnych zmian przez wielu użytkowników. Wygląd podstawowego ekranu z którym współpracuje użytkownik przedstawia rysunek 4. Aplikacja edytora umożliwia pracę z metadanymi zgromadzonymi w jednym węźle. Chroni to metadane przed przypadkowymi zmianami ze strony personelu innych węzłów.

Aplikacja dostępu do metadanych. Wygląd podstawowego ekranu przeglądarki metadanych przedstawia rysunek 5. Aplikacja umożliwia przeglądanie metadanych zorganizowanych w strukturę drzewiastą. Infrastruktura jest przygotowana do przechowywania metadanych węzłów zorganizowanych hierarchicznie. Sprzyja temu zastosowanie architektury rozproszonej.



Rys. 4. Główny ekran edytora metadanych



Rys. 5. Główny ekran przeglądarki metadanych

go katalogu metadanych z wbudowaną bramą. Niektóre elementy drzewa metadanych stanowią bramy do węzłów umieszczonych niżej w hierarchii. Tym sposobem użytkownik może trawersować kompletne drzewo metadanych pokonując granice serwerów. Aplikacja ta oferuje również użytkownikowi możliwość wyrafinowanego wyszukiwania metadanych. Metadane są podzielone na pakiety tematyczne zgodnie z zaleceniami normy ISO 19115 również w warstwie prezentacji.

W wypadku wyszukania odpowiedniego elementu metadanych, który opisuje dane mające reprezentację numeryczną przeglądarka metadanych umożliwia automatyczne przejście do aplikacji prezentowania danych.

Podsumowanie

Podsystem metadanych w RSIP jest zgodny z normą ISO 19115. Architektura podsystemu jest otwarta. Przyjęty model katalogu metadanych może być wykorzystany wprost (jak obecnie) lub może stać się częścią modelu z bramą wyniesioną poza struktury metadanych. Nie wymaga to żadnej przebudowy oprogramowania, gdyż w takim środowisku po prostu nie wystąpią elementy metadanych stanowiące bramy do innych węzłów.

Norma ISO 19115 została przyjęta jako podstawa budowania krajowych i regionalnych systemów metadanych w ramach europejskiego projektu INSPIRE. Katalogi metadanych winny zostać wyniesione poza same serwery na wzór *clearinghouse*. Podsystem metadanych zbudowany w ramach RSIP nie obejmuje takiego rozwiązania ze względu na ograniczenie projektu do pojedynczego województwa. Jednak po nawiązaniu współpracy ponadregionalnej sytuacja dojrzeje do budowania katalogów serwerów metadanych. Obecny podsystem – dzięki swej elastyczności – może stać się częścią większego systemu metadanych przez zarejestrowanie go w odpowiednim katalogu. Wyposażenie serwera metadanych w implementację protokołów wysokiego poziomu służących wymianie informacji, np. Z39.50 lub SOAP, również nie będzie trudne dzięki otwartej architekturze podsystemu i zastosowaniu wzorców architektonicznych.

Summary

Research on metadata began in 1980s. It was caused by an enormous growth of accumulated spatial data. Metadata contribute to put these huge data resources in order, providing their formalised documentation. Intensive efforts in many countries led to creation of interesting implementations of metadata systems and to drawing up an international standard ISO 19115. In Poland, the Institute of Spatial and Cadastral Systems began works connected with a Regional Spatial Information System. The architecture of the metadata subsystem of Regional Spatial Information System was developed on the basis of modern designing methods and fitted into the architecture of the whole system. The metadata subsystem is flexible and meets ISO 19115 standard, also with respect to specific profiles and extensions. It also takes into account recommendations of INSPIRE.

Edward Liszczyk
FORMIDA Laboratorium Informatyki
Gliwice
tel. (032) 234 13 71, e-mail: eliszczyk@op.pl