



**POLSKIE
TOWARZYSTWO
INFORMACJI
PRZESTRZENNEJ**

ROCZNIKI 2010

GEOMATYKI

Modelowanie danych przestrzennych

**Tom VIII
Zeszyt 4(40)
Warszawa**

Wprowadzenie

Niniejszy tom specjalny *Roczników Geomatyki* jest przeznaczony dla uczestników warsztatów nt. „Modelowania danych przestrzennych” zorganizowanych przez Instytut Geodezji i Kartografii, w ramach XX Konferencji Polskiego Towarzystwa Informacji Przestrzennej z cyklu GEOINFORMACJA W POLSCE. Zawiera on opis omawianych na kursie zagadnień związanych z aspektami teoretycznymi i praktycznymi modelowania danych przestrzennych, ze szczególnym zwróceniem uwagi na harmonizację i integrację tworzonych zbiorów danych przestrzennych.

Warsztaty są przeznaczone dla wszystkich tych, którzy chcą poznać lub pogłębić wiedzę teoretyczną oraz umiejętności praktyczne w zakresie modelowania danych przestrzennych zgodnie ze znormalizowanymi metodami. Ważnym elementem modelowania danych jest ich przekształcanie z jednych schematów aplikacyjnych na inne; warsztaty są, zatem dedykowane także tym osobom, które w pracy zawodowej staną przed koniecznością przekształcania zbiorów danych przestrzennych.

Niniejszy tom *Roczników Geomatyki* zawiera podstawy teoretyczne zagadnień poruszanych na warsztatach i stanowi uzupełnienie przykładów demonstrowanych w trakcie warsztatów. Autorami poszczególnych rozdziałów tomu i zajęć warsztatowych są specjaliści, z dużym doświadczeniem naukowym i dydaktycznym w zakresie modelowania geoinformacyjnego.

W kolejnych rozdziałach zeszytu zostały opisane aspekty teoretyczne i praktyczne modelowania danych przestrzennych szczególnie istotne z punktu widzenia budowania infrastruktury informacji przestrzennej i wdrażania dyrektywy INSPIRE w Polsce. Omówiono w nich znormalizowane zasady modelowania danych, ze szczególnym uwzględnieniem schematu pojęciowego i aplikacyjnego (rozdział 1), następnie przedstawiono uwarunkowania procesu przekształcania polskich zbiorów danych do wymagań specyfikacji opracowanych w ramach przepisów implementacyjnych dyrektywy INSPIRE (rozdział 2). Ponadto omówiono podstawowe różnice w zakresie technologicznym pomiędzy strukturami i formami danych polskich i danych zgodnych ze specyfikacjami INSPIRE. Ostatnim zagadnieniem poruszonym w ramach warsztatów i opisanym w rozdziale 3 są zasady i metodyka oceny jakości danych przestrzennych.

Uczestnicy warsztatów zdobędą podstawową wiedzę z zakresu przekształcania danych przestrzennych pomiędzy różnymi modelami danych, metod, technologii oraz narzędzi wykorzystywanych w procesie przekształcania danych. Przedstawiane rozwiązania oparte będą na doświadczeniu wynikającym z realizacji kursów doszkalających w zakresie *Modelowania pojęciowego w projektowaniu i implementacji systemów geoinformacyjnych* realizowanych przez Instytut Geodezji i Kartografii oraz na pracach eksperymentalnych prowadzonych w Laboratorium Modelowania i Geomatyki Uniwersytetu Warszawskiego, głównie przez dr hab. Janusza Michalaka.

Marek Baranowski
Elżbieta Bielecka

Rozdział 1. Modelowanie informacji geograficznej: Podstawy

Chapter 1. Geoinformation modeling: Fundamentals

prof. dr hab. inż. Wojciech Pachelski

Wydział Inżynierii Lądowej i Geodezji, Wojskowa Akademia Techniczna

WojciechPachelski@aster.pl

Rozdział 2. Transformacja polskich danych przestrzennych do modeli INSPIRE

Chapter 2. Transformation of Polish spatial data to INSPIRE models

dr hab. Janusz Michalak

Wydział Geologii, Uniwersytet Warszawski

J.Michalak@uw.edu.pl

<http://netgis.geo.uw.edu.pl>

Rozdział 3. Zasady oceny jakości danych przestrzennych oraz ich zastosowanie do oceny jakości danych gromadzonych w TBD

Chapter 3. Principles of spatial data quality assessment and their use for quality assessment of data stored in TBD

dr hab. inż. Elżbieta Bielecka, prof. WAT

Instytut Geodezji i Kartografii

Wydział Inżynierii Lądowej i Geodezji, Wojskowa Akademia Techniczna

elzbieta.bielecka@igik.edu.pl

MODELOWANIE DANYCH PRZESTRZENNYCH

Słowa kluczowe dla rozdziału 1: modelowanie pojęciowe, UML, schemat aplikacyjny, reguły budowy schematów aplikacyjnych

Słowa kluczowe dla rozdziału 2: model INSPIRE, transformacja danych, XSLT, GML, XML

Słowa kluczowe dla rozdziału 3: dane przestrzenne, jakość danych, normy ISO, TBD, INSPIRE

Streszczenie

Niniejszy zeszyt specjalny *Roczników Geomatyki* jest poświęcony wybranym zagadnieniom modelowania danych przestrzennych, które były omawiane podczas warsztatów nt. „Modelowania danych przestrzennych”, zorganizowanych w ramach XX Konferencji Polskiego Towarzystwa Informatyki Przestrzennej z cyklu GEOINFORMACJA W POLSCE. W kolejnych rozdziałach zeszytu zostały opisane aspekty teoretyczne i praktyczne modelowania danych przestrzennych szczególnie istotne z punktu widzenia budowy infrastruktury informacji przestrzennej i wdrażania dyrektywy INSPIRE w Polsce.

Zeszyt rozpoczyna rozdział *Modelowanie informacji geograficznej: Podstawy*, stanowiący teoretyczny wstęp do poruszanej problematyki. Modelowanie informacji jest najważniejszym etapem w budowie systemu informacyjnego, także w przypadku systemu informacji geograficznej. Modelowanie to polega na ścisłym, kompletnym i sformalizowanym opisie kategorii obiektów wyróżnionych w danym obszarze przedmiotowym rzeczywistości. Ma ono szczególne znaczenie zwłaszcza w obszarze informacji geograficznej, gdzie od poprawności i kompatybilności przyjętych rozwiązań zależy efektywność współdziałania licznych, wielorakich tematycznie, różnorodnych narzędziowo oraz rozproszonych instytucjonalnie i regionalnie infrastruktur danych przestrzennych, będących przedmiotem wielu programów międzynarodowych. W tym celu stosowane są zaawansowane języki formalne, jak UML, oraz zespół metod, pojęć i narzędzi ujętych w normy międzynarodowe ISO. W artykule zostały omówione także podstawowe elementy notacyjne diagramów klas UML, reguły budowy schematów aplikacyjnych, integracja budowanego modelu ze schematami znormalizowanymi opisu położenia, geometrii i topologii oraz jakości i metadanych. Przedstawiono także przykłady wykorzystania omówionych reguł do opisu struktur informacyjnych, zgodnych ze standardami technicznymi wydanymi przez Głównego Geodetę Kraju.

W rozdziale drugim zostały opisane zagadnienia harmonizacji i transformacji polskich danych przestrzennych do modeli INSPIRE, zarówno od strony podstaw teoretycznych, jak i konkretnych przykładów zapisania danych gromadzonych w polskich zasobach zgodnie z wymaganiami specyfikacji implementacyjnych dyrektywy INSPIRE. Podstawy do praktycznego rozwiązywania problemów z zakresu harmonizacji modeli danych przestrzennych i transformacji danych pomiędzy różnymi modelami zostały opisane w dokumentach OGC i Komitetu ISO/TC211. Podstawą metodologiczną, zalecaną przez obie organizacje, są modele pojęciowe danych zapisane w języku UML, na podstawie których można opracowywać struktury baz danych i schematy XML specyfikujące języki aplikacyjne zapisu danych w plikach dla przechowywania danych w repozytoriach lub przesyłania ich między różnymi systemami. Podstawowym językiem z rodziny XML dla danych przestrzennych jest GML, a praktyczne jego wykorzystania w różnych dziedzinach są realizowane przy pomocy jego aplikacji. Do tej kategorii zalicza się także specyfikacje danych dla poszczególnych tematów INSPIRE.

W kolejnych częściach rozdziału omówiono różne aspekty przedstawianej problematyki – od podstawowych pojęć z zakresu transformacji danych przestrzennych i języków znacznikowych, poprzez przedstawienie głównych różnic pomiędzy formatem zapisu danych a językiem służącym do tego zapisu i problemów implementacji języka GML, aż do przedstawienia podstaw transformacji zbiorów XML przy pomocy technologii XSLT. Znaczna część rozdziału poświęcona jest aspektom praktycznym, w tym schematom aplikacyjnym danych INSPIRE, oprogramowaniu dedykowanemu transformacji, analizie próbek polskich zbiorów danych przestrzennych poddanych testom i praktycznym metodom transformacji z propozycją ogólnego algorytmu opisującego poszczególne jej fazy.

Rozdział kończy się podsumowaniem, które jest próbą wyciągnięcia bardziej ogólnych wniosków i uwag wynikających z opisanych w nim prac badawczych. W przedstawionych pracach testowane były tylko małe próbki danych przestrzennych, głównie z zakresu pierwszej grupy tematycznej INSPIRE, jednak uzyskane wyniki mogą być ekstrapolowane na znacznie szerszą skalę. Przedstawione wyniki prac studialnych i testowych nad rozwojem technologii takiej transformacji, a także wnioski ogólne i praktyczne, zdaniem Autora będą pomocne w przyszłych pracach prowadzonych w skali pełnych zasobów dziedzinowych i w skali całego kraju.

Rozdział trzeci dotyczy zasad oceny jakości danych przestrzennych. Wobec powszechnego dostępu do danych przestrzennych znajomość ich jakości staje się zagadnieniem kluczowym zarówno dla użytkowników danych, jak też instytucji je udostępniających. Jakość danych nabiera szczególnego znaczenia, gdy informacje uzyskane na ich podstawie służą do podejmowania decyzji. Wiadomo, że decyzje podejmuje się często bazując na informacjach niepewnych, jednak każdorazowo decydent musi mieć tego pełną świadomość. W procesie podejmowania decyzji ocena dokładności i wiarygodności danych (informacji) ma kluczowe znaczenie. W opracowaniu przedstawiono podstawy oceny jakości danych zgodnie z zasadami podanymi w normach ISO serii 19100 oraz ocenę zgodności danych zgromadzonych w TBD z Wytocznymi TBD, zwaną oceną jakości „producenta”. W zakresie oceny danych TBD przyporządkowano elementom kontroli jakości wg ISO 19113 i specyfikacji technicznych INSPIRE odpowiednie zestawy kontroli automatycznych TBD, przeprowadzonych zgodnie z wytycznymi technicznymi i koncepcją systemu kontroli TBD. Ponadto dokonano interpretacji wykrytych zdarzeń, a także przedstawiono dodatkowe procedury kontrolne, konieczne do całościowej oceny jakości zbioru zgodnie z wymaganiami przepisów implementacyjnych dyrektywy INSPIRE.

SPATIAL DATA MODELING

Keywords for Chapter 1: conceptual modeling, UML, application schema, rules for application schema

Keywords for Chapter 2: INSPIRE model, data transformation, XSLT, GML, XML

Keywords for Chapter 3: spatial data, data quality, ISO standards, TBD, INSPIRE

Abstract

This special issue of *Annals of Geomatics* is devoted to selected problems related to spatial data modeling, discussed at the workshop „Spatial data modeling” organized within the framework of the XX Conference GEOINFORMATION IN POLAND of the Polish Association for Spatial Information. In four chapters, theoretical and practical aspects of spatial data modeling are discussed, particularly important from the point of view of creation of the spatial information infrastructure in Poland and implementation of the INSPIRE Directive.

Chapter 1 *Geoinformation modeling: Fundamentals* constitutes a theoretical introduction to the problems discussed. Information modeling is the most important stage in the construction of a geoinformation system. Modeling consists in strict, complete and formalized description of object categories distinguished in a given subject area of reality. This is particularly important for geoinformation, where correctness and compatibility of the solutions is essential for the efficiency of cooperation of numerous institutionally and regionally dispersed spatial information infrastructures covering several themes and using a variety of tools. In addition, SII are engaged in many international programs. For this purpose, advanced formal languages are used such as UML and a set of methods, concepts and tools contained in the international ISO standards. The chapter also discusses basic notation elements of UML diagrams, construction rules for application schemas, integration of the model constructed with normalized schemas describing location, geometry and topology as well as quality and metadata. The paper also contains examples of utilization of the rules discussed for description of information structures compatible with technical standards issued by the General Surveyor of Poland.

In Chapter 2 problems connected with harmonization and transformation of Polish spatial data to INSPIRE models are described both from theoretical point of view and as concrete examples of recorded data stored in Polish resources in accordance with the requirements of implementation specification of the INSPIRE Directive. The basis for practical solution of the problems related to harmonization of spatial data models and data transformation between different models are described in OGC and Committee ISO/TC211 documents. Conceptual data models written in UML language constitute the metodological basis recommended by both organizations and on this basis structures of data bases may be developed as well as XML schemas specifying application languages for description of data in files for data storage in repositories or for transferring them between different systems. The basic language from XML family for spatial data is GML and its practical use in various areas is realized by means of its application. Data specifications for individual INSPIRE themes are also included to this category.

Further in the Chapter various aspects of the problems presented are discussed – from basic notions in the area of spatial data transformation and marker languages, through basic differences

between the format of data records and the language serving this recording and the problems connected with implementation of GML language through presentation of the basis for transformation of XML sets by means of XSLT technology. A large part of the Chapter is devoted to practical aspects, including application schemas of INSPIRE data, software dedicated to transformation, analysis of samples of Polish spatial data sets subjected to tests and practical methods of transformation with proposed general algorithm describing its individual stages.

The Chapter closes with a summary striving to draw more general conclusions and remarks resulting from the research work described. During the research work not only small samples of spatial data were tested, mostly from the first INSPIRE thematic group. However, the results obtained may be extrapolated on much larger scale. The presented results of study and test works on development of technology of such a transformation as well as general and practical conclusions will be helpful – in the author’s opinion – in future works conducted on the scale of full domain resources and on the scale of the whole country.

Chapter 3 refers to the principles of spatial data quality assessment. Taking into account that spatial data are generally accessible, the awareness of their quality is of key importance both for the users of these data and for the institutions making them available. Data quality is of special importance when the information obtained serves as the basis for decision making. It is well known that decisions are often taken on the basis of uncertain information, but in every case the decision maker must be fully aware of this. In the decision making process accuracy and trustworthiness of data is of key importance. The basis for data quality assessment are presented in this work in accordance with the principles contained in ISO standards series 19100 and assessment of compatibility of the data stored in TBD with TBD Guidelines, called “manufacturer’s” quality assessment. As regards assessment of TBD data, appropriate sets of automatic TBD controls conducted in accordance with technical guidelines and the concept of TBD control system were subordinated to the elements of quality control according to ISO 19113 and INSPIRE technical specification. Moreover, discovered events were interpreted and additional control procedures were presented necessary for comprehensive quality assessment of the set in accordance with the requirements of the implementation provisions of the INSPIRE Directive.

Spis treści

1. Modelowanie informacji geograficznej: Podstawy	13
1.1. Wstęp: pojęcia podstawowe	13
1.2. Proces modelowania informacji geograficznej	15
1.3. Diagramy klas UML: elementy notacyjne	17
1.4. Reguły budowy schematu aplikacyjnego	21
1.5. Schemat przestrzenny: opisywanie geometrii, topologii i położenia	23
1.6. Integracja schematu aplikacyjnego ze schematami znormalizowanymi	25
1.7. Przykład modelu pojęciowego	26
2. Transformacja polskich danych przestrzennych do modeli INSPIRE	27
2.1. Podstawowe pojęcia	29
2.2. Różnice pomiędzy formatem i językiem	32
2.3. Problemy implementacji języka GML	36
2.4. Specyfikacje danych INSPIRE	39
2.5. Oprogramowanie dedykowane transformacji	41
2.6. Testowanie metod transformacji	43
2.7. Zbiory danych testowych	46
2.8. Transformacja zbiorów danych w językach aplikacyjnych XML	47
2.9. Podsumowanie	49
3. Zasady oceny jakości danych przestrzennych oraz ich zastosowanie do oceny jakości danych gromadzonych w TBD	53
4.1. Wprowadzenie	53
4.2. Ocena jakości danych przestrzennych według norm ISO serii 19 100	53
4.3. Procedura określania jakości danych	56
4.4. Założenia kontroli jakości danych gromadzonych w TBD	57
4.5. Ocena jakości danych gromadzonych w zasobie podstawowym TBD	59
4.5.1. Założenia wstępne	59
4.5.2. Metodyka oceny jakości danych zasobu podstawowego TBD i otrzymane wyniki	59
4.5.3. Podsumowanie oceny jakości danych i dyskusja	63
4.5. Znaczenie oceny jakości danych	65