



**POLSKIE
TOWARZYSTWO
INFORMACJI
PRZESTRZENNEJ**

ROCZNIKI 2010

GEOMATYKI

Modelowanie danych przestrzennych

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

Wojciech Pachelski

1. MODELOWANIE INFORMACJI GEOGRAFICZNEJ: PODSTAWY

1.1. Wstęp: pojęcia podstawowe

Niewątpliwie najbardziej podstawowym pojęciem w tematyce ujętej w tytule jest *informacja*, którego istota jest nieuchwytna i trudna do zdefiniowania tak, że często zadowalamy się jej intuicyjnie domniemywanym sensem. W książce Schencka i Wilsona (1994) podane są następujące przykłady różnych form informacji:

- przekazana lub odebrana wiedza;
- wiedza uzyskana z badań, studiów lub instruktażu;
- atrybut właściwy uporządkowaniu dowolnych elementów i przekazywany poprzez to uporządkowanie (np. geny);
- informacja o faktach wykorzystywana jako podstawa rozumowania, dyskusji lub kalkulacji.

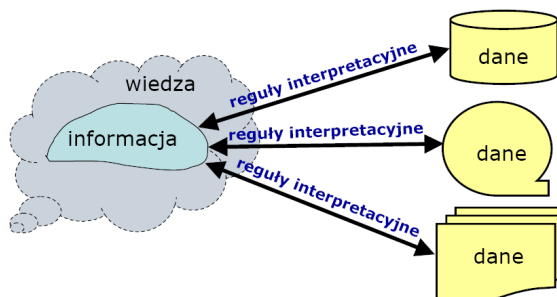
Występujące w tych przykładach zbliżone pojęcia takie, jak *wiedza*, *komunikowanie* i *dane*, prowadzą do następujących definicji zawartych w normie PN-EN-ISO 19101 (ISO, 2002):

Informacja jest to wiedza dotycząca obiektów takich jak fakty, zdarzenia, rzeczy, procesy, pojęcia, koncepcje, która w określonym kontekście ma konkretne znaczenie.

Dane stanowią interpretowalną i sformalizowaną reprezentację informacji, stosowną do jej komunikowania, interpretowania lub/i przetwarzania.

Dane, czyli zespół znaków, symboli i różnego rodzaju sygnałów, powstaje przez zastosowanie do informacji określonych reguł, tzw. **reguł interpretacyjnych** (np. reguł języka polskiego, kodu ASCII, formatu obrazu JPG, formatu dźwięku mp3 itp.). Aby z kolei zinterpretować przekazane dane i odzyskać zawartą w nich informację, odbiorca danych musi zastosować do nich te same reguły interpretacyjne (lecz „w odwrotnym kierunku”), jakie były wcześniej użyte do ich utworzenia.

O ile zatem istotą informacji jest jej znaczenie i sens, czyli semantyka, to dane są formą symboliczną i syntaktyczną informacji. Jawne bądź ukryte reguły interpretacyjne są przy tym czynnikiem niezbędnym dla współistnienia informacji i danych. Jedna i ta sama informacja może być przedstawiona jako różne dane, utworzone za pomocą odmiennych reguł interpretacyjnych. Jednakże zależność odwrotna, tj. możliwość wieloznacznej interpretacji danych, byłaby oczywiście przypadkiem zaprzeczającym istocie informacji. Omówione powyżej zależności pomiędzy informacją, danymi i regułami interpretacyjnymi są przedstawione na rysunku 1.1.



Rys. 1. Informacja i dane oraz reguły interpretacyjne

Z kolei informacja geograficzna wyróżnia się tym, że pewne obiekty, których dotyczy, posiadają cechy przestrzenne – jedno-, dwu- lub trójwymiarowe – takie, jak kształty i wzajemne relacje obiektów opisywane w kategoriach geometrii i topologii, wymiary oraz położenie odniesione do powierzchni Ziemi. To właśnie te cechy decydują o zakwalifikowaniu danej informacji do tej właśnie kategorii. Mają one również istotne konsekwencje dla problemów

przechowywania, przetwarzania, udostępniania i innych, co stwarza konieczność stosowania specyficznych modeli danych i algorytmów do ich przetwarzania. Definicja informacji geograficznej i danych geograficznych zawarta jest, jak poprzednio, w normie PN-EN-ISO 19101 (ISO, 2002), a mianowicie:

Informacja geograficzna jest to informacja dotycząca zjawisk jawnie bądź niejawnie powiązanych z położeniem odniesionym do Ziemi.

Dane geograficzne są to dane z jawnym bądź niejawnym odniesieniem do położenia na Ziemi.

W terminologii polskiej, w literaturze i praktyce geoinformacyjnej, są często stosowane terminy *informacja przestrzenna* i *dane przestrzenne*, które należy traktować jako odpowiedniki normatywnych zwrotów *informacja geograficzna* i *dane geograficzne*.

Nieodzowną cechą informacji jest możliwość jej udostępnienia, przekazania, odebrania czy też wymiany – ogólnie mówiąc – jej komunikowania, co jest koniecznym warunkiem współdziałania (*interoperacyjności*¹) poszczególnych składników tzw. *infrastruktury danych przestrzennych*. Dla spełnienia tego warunku oczywiście niezbędna jest postać informacji jako danych wraz z właściwym zestawem reguł interpretacyjnych.

Model informacyjny jest to sformalizowany opis obiektów, idei, faktów i procesów co do ich **typów**, który łącznie tworzy model, czyli stanowi charakterystykę, interesującego nas fragmentu świata rzeczywistego i który dostarcza jawnego zestawu **reguł interpretacyjnych** (Schenk i Wilson, 1994).

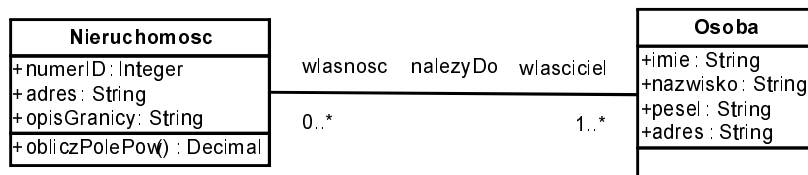
Przez **typ obiektu** w powyższej definicji należy rozumieć ogólną postać obiektu (także idei, faktu, itp.), nie zaś obiekt konkretny (ideę, fakt,...). Na przykład, jeśli takim typem obiektu jest budynek, mamy na myśli kategorię pojęciową „budynek” w ogóle, czyli każdą konstrukcję drewnianą lub murowaną, która posiada pewną liczbę kondygnacji i zadaszenie, ma jakieś przeznaczenie, adres, itp., czyli posiada cechy właściwe budynkowi. Typ obiektu jest zatem ogólną abstrakcją, którą należy odróżnić od konkretnego **przypadku** (*egzemplarza* lub *instancji*)² obiektu, jakim może być, wśród wielu możliwych takich przypadków, na przykład konkretny budynek czteropiętrowy, pokryty dachówką, mieszkalny, położony w Olsztynie przy ul. Jarockiej 76b. Mamy więc do czynienia z poziomem **typów obiektów** oraz z poziomem **przypadków obiektów**. Poziom typu i poziom przypadku rozciąga się na inne twory modelowe, jak atrybuty, związki, operacje. Ogólnie mówimy zatem o **poziomie typów**

¹ Terminy *współdziałanie* i *interoperacyjność* traktowane są jako synonimy.

² W terminologii angielskiej *instance*, stąd spotyka się tłumaczenie – *instancja*. W języku polskim najbliższym znaczeniowo jest słowo *przypadek*.

danych i poziomie przypadków danych. Model pojęciowy budowany jest z reguły na poziomie typów danych.

Kolejną cechą modelu informacyjnego jest to, że posiada on *sformalizowany opis*. Jest on zbudowany ze skończonej liczby elementarnych pojęć, z których każde ma dokładnie określone znaczenie. Pojęcia elementarne są łączone według określonych reguł i zasad w bardziej złożone konstrukcje, tworząc z kolei byty również posiadające sprecyzowane znaczenia. Jednym z najnowszych formalizmów stosowanych do opisu modelu informacyjnego, zalecanym przez normy międzynarodowe i europejskie, jest język UML³, w którym do pojęć elementarnych należą m.in.: *klasa* jako reprezentacja określonego typu obiektu, *atrybut* jako reprezentacja określonej cechy obiektu, *operacja* jako reprezentacja określonego działania przypisywanego obiektowi oraz *związek* jako reprezentacja określonej zależności pomiędzy obiektami. Za pomocą tych pojęć elementarnych można zbudować, na przykład model informacyjny katastru, w skład którego, wśród wielu innych składników, może wchodzić m.in. zależność pomiędzy nieruchomością a osobą w roli właściciela. W przykładzie tym typ obiektu *Nieruchomosc* może być reprezentowany jako klasa o tejże nazwie, której przypisane są atrybuty takie, jak m.in. *numerId*, *adres*, *opisGranic*⁴, jak też operacje – np. *obliczPolePow*. Podobnie, typ obiektu *Osoba* może być reprezentowany jako klasa o atrybutach *imię*, *nazwisko*, *pesel*, *adres*, itp. Pomiędzy nieruchomością a osobą istnieje związek: *Nieruchomosc* – *należyDo* – *Osoba*. Omawiany fragment modelu informacyjnego katastru można zapisać w notacji UML jako tzw. diagram klas, przedstawiony na rysunku 1.2.



Rys. 1.2. Przykładowy diagram klas UML jako fragment modelu informacyjnego katastru

1.2. Proces modelowania informacji geograficznej

Modelowanie informacji jest rozumiane na ogół jako proces budowy systemu informacyjnego, na który składają się następujące wzajemnie przenikające się etapy:

- **modelowanie pojęciowe**, w którym analizowane są w kategoriach ogólnych, niezależnych od środowiska komputerowego, typy obiektów, ich właściwości, operacje i powiązania z innymi obiektami;
- **modelowanie logiczne**, w którym model pojęciowy jest uzupełniany o ogólne, wspólne dla różnych typów platform narzędziowych, aspekty cyfrowej reprezentacji danych, jak m.in. przypisanie danym stosownych typów wartości (liczbowych, logicznych, znakowych, itp.);
- **modelowanie fizyczne**, nazywane też modelowaniem danych, w wyniku którego powstaje, zwykle w sposób zautomatyzowany, baza danych dostosowana do konkretnej platformy komputerowej, obejmująca pliki, rekordy, kody itp. konstrukcje właściwe tej platformie.

³ UML – *Unified Modeling Language*.

⁴ Sposób nadawania nazw konstrukcjom UML jest zgodny z normami ISO.

Spośród powyższych etapów najważniejszym jest modelowanie pojęciowe. W etapie tym, abstrahującym od konkretnej platformy sprzętowo-programowej, zawarta jest bowiem w kategoriach ogólnych zasadnicza koncepcja i projekt budowy systemu informacyjnego, które mogą z kolei podlegać realizacji (implementacji) na dowolnej platformie (spełniającej określone wymagania techniczne). Występuje tu daleko idąca analogia modelowania pojęciowego do projektowania dowolnej konstrukcji inżynierskiej (dróg, mostów, maszyn, komputerów, itp.): w obu przypadkach przedmiotem jest twórcza myśl inżynierska ujmowana w specyficzny formalizm, spełniająca (lub nie) określone kryteria jakościowe, ekonomiczne i inne, mogąca podlegać realizacji, praktycznej weryfikacji, itp.

Model pojęciowy jest to model definiujący pojęcia z obszaru będącego przedmiotem zainteresowania (PN-EN-ISO 19101:2005 Informacja geograficzna – Model odniesienia).

Model pojęciowy powstaje w wyniku szczegółowej analizy określonej dziedziny przedmiotowej (tematycznej). Modelowanie pojęciowe jest procesem tworzenia abstrakcyjnego opisu pewnego fragmentu świata rzeczywistego wraz ze zbiorem odnośnych pojęć i przy wykorzystaniu określonych reguł (zob. rys. 1.3). Przykładem takiego fragmentu świata rzeczywistego może być zbiór obiektów, jak ciekły wodne, jeziora, wyspy i inne. Dla opisanie ich kształtów należy wykorzystać pewne konstrukcje geometryczne, jak punkty, linie, powierzchnie, które stanowią tutaj ten zbiór pojęć dodatkowych. Podobnie, dla opisanie ich położenia czy innych cech należy wykorzystać „predefiniowane” reguły i zasady. Model taki może istnieć jedynie w umysłach jego twórców, może też być zapisany słownie, przeważnie w sposób nieprecyzyjny, i w tej postaci może być też upowszechniany. Dla precyzyjnego i jednoznacznego zapisu modelu niezbędny jest zatem stosowny język formalny, jak np. UML czy EXPRESS, który zawiera właściwe środki składniowe i znaczeniowe oraz jest przetwarzalny komputerowo. Model pojęciowy zapisany w takim języku nosi nazwę *schematu pojęciowego* i może być w sposób zautomatyzowany, tj. za pomocą odpowiednich narzędzi komputerowych, przetworzony do postaci modelu logicznego, a następnie „zaimplementowany” jako model fizyczny na wybranej platformie komputerowej. W ten sposób jeden i ten sam model pojęciowy może być podstawą zgodnych implementacji na różnych platformach, co jest jednym z czynników zapewniających niezbędne współdziałanie różnych systemów informacyjnych. Warunkiem współdziałania takich systemów jest zapewnienie jednolitych reguł budowy modeli i ich opisów w postaci schematów. Postępowanie mające na celu zapewnienie zgodności pomiędzy różnymi schematami pojęciowymi jest znane jako *integracja modelu* i jest przedmiotem normy PN-EN-ISO 19109 (ISO, 2005a).

Istotne znaczenie dla modelu pojęciowego ma tzw. ogólny model obiektów GFM (*General Feature Model*). Model ten zawiera pojęcia niezbędne do klasyfikacji obiektów będących przedmiotem modelowania pojęciowego oraz opisuje strukturę katalogu obiektów. Struktura ta jest przedmiotem normy PN-EN-ISO 19110 (ISO, 2005b).

Z punktu widzenia budowy systemu informacji geograficznej rolę modelowania pojęciowego przedstawia rysunek 1.4. Rysunek ten wyróżnia, od najbardziej abstrakcyjnych do coraz to bardziej szczegółowych, kilka poziomów realizacyjnych systemu informacyjnego:

- I. od poziomu podstaw z zakresu teorii poznania i ontologii, dających pogląd na istotę i charakter obiektów, procesów i zjawisk będących przedmiotem opisu w kategoriach modelu pojęciowego; poziom ten nie jest rozważany w ramach niniejszego opracowania;
- II. poprzez poziom metodologii, w ramach której formułowane są ogólne i uniwersalne, abstrahujące od środowisk instytucjonalnych, przedmiotowych i komputerowych, pojęcia, teorie, prawa, metody i reguły budowy modeli informacyjnych;

III. poziom technologii rozumianej jako aplikacje narzędziowe GIS gotowe do zainstalowania w poszczególnych takich środowiskach;

IV. do poziomu bezpośrednich zastosowań w określonych obszarach przedmiotowych.

Rozpatrując powyższą systematykę w kierunku odwrotnym – od poziomu IV, tj. od zainstalowanych i działających w odpowiednich ośrodkach wykonawczych systemów GIS wraz z napełnionymi treścią bazami danych, zauważmy, że to właśnie na tym poziomie ma nastąpić wymiana danych, która jest najbardziej istotnym czynnikiem współdziałania GIS. Zapewnienie współdziałania na tym poziomie musiałyby oznaczać odgórne ustalenie jednego z ośrodków jako centrum przyjmowania i rozdziału danych w sposób podobny do działania np. urzędów pocztowych, co byłoby rozwiązaniem z gruntu niepraktycznym, a nawet utopijskim. Podobnie nieracjonalne byłoby poszukiwanie rozwiązania na poziomie III – technologicznym, gdzie uprzywilejowany charakter musiałby być zadekretowany jednej wybranej aplikacji narzędziowej wraz z żądaniem, by pozostałe aplikacje akceptowały struktury, formaty, kody itp. danych zgodne ze strukturami, formatami i kodami tej aplikacji wybranej. Łatwo zauważyć, że oba powyższe rozwiązania, poza oczywistą nieracjonalnością, miałyby charakter dekretowy, przeciwdziałający wszelkim modyfikacjom i modernizacjom raz przyjętych rozwiązań, a tym samym naturalnemu rozwojowi infrastruktur informacji przestrzennej.

Poziom III – technologiczny i poziom IV – zastosowania reprezentują modele danych – logiczny i fizyczny – a więc są zależne od tych środowisk, tym samym nie mogą być podstawą dla rozwiązań ogólnych, uniwersalnych. W tej sytuacji racjonalny, systemowy i ogólny charakter, zapewniający współdziałanie, mają jedynie rozwiązania metodologiczne, tj. na poziomie II według powyższej systematyki. Rozwiązania te, oparte na modelowaniu pojęciowym, dotyczą bowiem struktur informacyjnych, które są niezależne od środowisk komputerowych.

Rozwiązania metodologiczne sprowadzają się nie tylko do zapewnienia, w formie modeli pojęciowych, wspólnych, ogólnych i uniwersalnych podstaw dla budowy oddzielnych zgodnych aplikacji – modeli fizycznych – w zróżnicowanych środowiskach komputerowych, przedmiotowych, instytucjonalnych i innych, lecz również do zapewnienia jednolitych podstaw teoretycznych, zasad, reguł, metod, itp., dla budowy i opisu tych modeli. Ten ostatni czynnik jest zapewniony przez zespół wzajemnie powiązanych norm metodycznych serii ISO 19100, które jako zaadaptowane normy europejskie (EN) są podstawą budowy europejskiej infrastruktury informacji przestrzennej w ramach projektu INSPIRE (2007), jako zaś Polskie Normy (PN) – są podobną podstawą dla budowy krajowej infrastruktury informacji przestrzennej.

1.3. Diagramy klas UML: elementy notacyjne

W myśl oficjalnej definicji twórców języka UML⁵ (Grady Booch, Ivar Jacobson, James Rumbaugh; *RATIONAL Software Corporation*; Booch i in., 2002) jest to *zunifikowany język modelowania, przeznaczony do specyfikowania, konstruowania, wizualizacji i dokumentowania elementów systemów*. Jest to język (notacja graficzna) modelowania pojęciowego informacji w postaci schematów pojęciowych. Umożliwia zautomatyzowane (komputerowe) przekształcanie tych schematów do zapisu w języku leksykalnym, jak np. w języku XML.

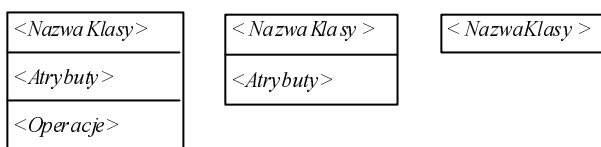
Dla celów modelowania informacji geograficznej według norm ISO serii 19100 sformułowano tzw. profil UML (ISO, 2005), który zawiera reguły stosowania UML w tej dziedzinie

⁵ *Unified Modeling Language* – zunifikowany język modelowania.

informacji geograficznej w sposób dostosowany do norm ISO serii 19100, zgodny z ogólną normą UML: ISO/IEC 19501 oraz spełniający pewne dodatkowe warunki budowy schematów pojęciowych informacji geograficznej.

Główną jednostką syntaktyczną tego języka jest tzw. diagram klas. Na schemat pojęciowy danego modelu składa się zwykle szereg tego typu diagramów. Podstawowe elementy notacji diagramów klas są omówione poniżej.

I. **Klasa.** Opis lub reprezentacja pewnej jednolitej kategorii obiektów, które są charakteryzowane za pomocą atrybutów tego samego typu, mają przypisane tego samego typu operacje i metody oraz pozostają w takich samych związkach z obiektami innych klas. W diagramach klas klasa jest przedstawiana w sposób pokazany na rysunku 1.5.



Rys. 1.5. Sposoby przedstawienia klasy w diagramie klas UML

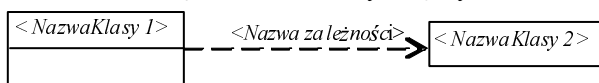
II. **Atrybut.** Każdy atrybut opisuje cechę lub właściwość obiektu danej klasy. W związku z tym atrybut przyjmuje wartość z określonej dziedziny wartości, czyli tzw. typu danych, np. wartość liczbową, tekstową, logiczną lub inną. W profilu UML dostępne są następujące typy danych:

- a) elementarne (*primitive*): Integer, Decimal, Real, Vector, CharacterString, Date, Time, DateTime, Boolean, Logical, Probability, Multiplicity;
- b) implementacyjne (*implementation*):
 - i. zbiorowe (*collection*): Set, Bag, Sequence, Dictionary;
 - ii. wyliczeniowe (*enumeration*): Enumeration, CodeList;
 - iii. representation: Record (RecordType), GenericName;
- c) pochodne (*derived*): Area, Length (Distance), Angle, Scale, Mtime, Volume, Velocity.

Pełny wykaz typów danych zawarty jest w dokumencie (ISO, 19103). Oprócz tego dozwolone są typy danych definiowane przez użytkownika.

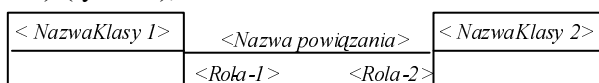
III. **Związek.** W UMLu dostępny jest szereg typów powiązań pomiędzy obiektami różnych klas, a mianowicie:

- a) **zależność:** związek znaczeniowy między dwoma elementami (rys. 1.6);



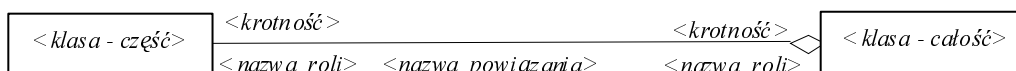
Rys. 1.6. Przedstawienie związku zależności w UML

- b) **powiązanie (asocjacja):** dowolny związek strukturalny pomiędzy obiektami, który jest istotny dla budowanego modelu (np. klient może wypożyczyć z biblioteki kilka książek) (rys. 1.7);



Rys. 1.7. Przedstawienie związku powiązania w UML

- c) **agregacja** (agregacja zwykła): związek zawierania się pomiędzy klasami: klasa jednego typu (całość – pojemnik) składa się z zestawu innych klas, które stanowią jego składniki (części); klasy-części mogą być składnikami innych agregacji, zaś ich czas życia nie jest ograniczony do czasu życia klasy-całości (zob. rys. 1.8); *<krotność>* oznacza przy tym liczbę powiązanych przypadków danej klasy: puste = 1, 0..1 – zero lub jeden, 0..* – zero lub wiele, 1..* – jeden lub wiele, *n* – konkretna liczba;



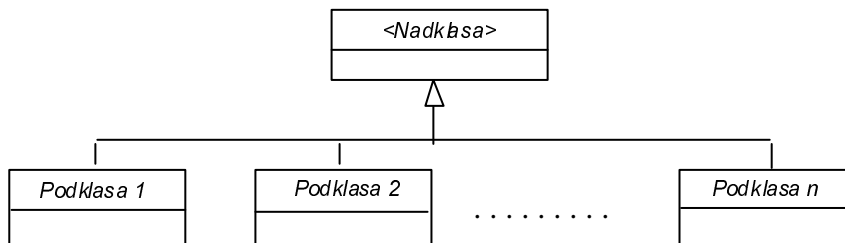
Rys. 1.8. Przedstawienie agregacji zwykłej w UML

- d) **kompozycja** (in. agregacja mocna lub całkowita): obiekty-części nie mogą być składnikami innych obiektów-całości, zaś ich czas życia jest ograniczony do czasu życia obiektu-całości (zob. rys. 1.9);



Rys. 1.9. Przedstawienie kompozycji w UML

- e) **uogólnienie** (in. dziedziczenie, generalizacja-specjalizacja): powiązanie, w którym jedna klasa (podtyp) dziedziczy własności z innej klasy (nadtypu); nadtyp posiada właściwości wspólne dla wszystkich swoich podtypów; podtyp dziedziczy wszystkie atrybuty, metody i powiązania nadtypu (zob. rys. 1.10);

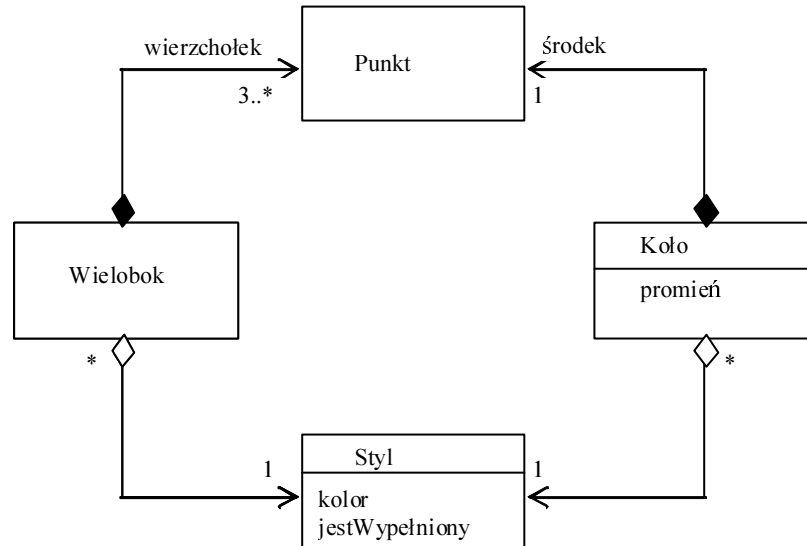


Rys. 1.10. Przedstawienie uogólnienia w UML

IV. **Operacja**: operacje dostarczają informacji o zmianach lub zachowaniach obiektów w ramach danej klasy na skutek działania czynników zewnętrznych, jak na przykład zamknięcie konkretnej drogi w porze zimowej, co powoduje brak połączenia z określonymi miejscowościami; wykonanie danej operacji dostarcza nowej wartości określonego typu.

V. **Ograniczenie**: dowolny tekst w pseudo-kodzie, języku formalnym lub języku programowania formułujący warunki i ograniczenia co do wartości atrybutów i operacji lub związków.

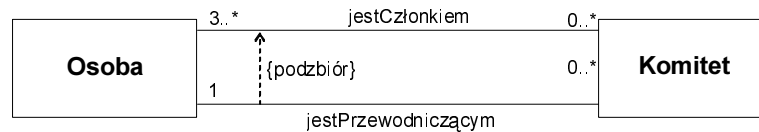
Przykład 1. Agregację zwykłą i kompozycję ilustruje diagram na rysunku 1.11.



Rys. 1.11. Przykład użycia związków agregacji i kompozycji UML

Dowolny przypadek (obiekt) klasy **Punkt** może należeć albo do obiektu klasy **Wielobok** (jako wierzchołek wieloboku), albo do obiektu klasy **Koło**, ale nie do obu jednocześnie. Z drugiej strony każdy obiekt klasy **Styl** może być współdzielony przez wiele obiektów klas **Wielokąt** i **Koło**. Usunięcie dowolnego obiektu każdej z tych klas powoduje usunięcie wszystkich stowarzyszonych z nim obiektów klasy **Punkt**, lecz żadnego stowarzyszonego z nim obiektu klasy **Styl**.

Przykład 2. Użycie związku zależności jako ograniczenia istniejących związków między klasami ilustruje diagram na rysunku 1.12.



Rys. 1.12. Przykład użycia związku zależności jako ograniczenia istniejących związków

Konkretna osoba może być członkiem lub przewodniczącym od zera do wielu komitetów. Konkretny komitet ma w swym składzie od trzech do wielu osób, jak również jedna konkretna osoba jest przewodniczącym komitetu. Zaznaczony związek zależności wyraża warunek, by przewodniczący komitetu rekrutował się spośród członków tegoż komitetu.

1.4. Reguły budowy schematu aplikacyjnego

Schemat aplikacyjny jest wynikiem modelowania pojęciowego, będącego najważniejszym składnikiem metodologii budowy infrastruktury informacji przestrzennej. Schemat ten jest sformalizowanym, jednoznacznym i kompletnym opisem struktur informacyjnych dla określonego zakresu tematycznego. Normą kluczową dla budowy schematu aplikacyjnego jest norma PN-EN-ISO 19109 *Reguły schematów aplikacyjnych* (ISO, 2005a). Jako formalizm zapisu schematu zaleca się dostosowany do potrzeb omawianej metodologii język UML według specyfikacji technicznej ISO/TS 19103: *Geographic information – Conceptual Schema Language* (ISO, 2005b).

Schemat aplikacyjny jest to *schemat pojęciowy* dla danych wykorzystywanych przez jedną lub więcej *aplikacji*⁶ (ISO, 2002; INSPIRE, 2008a; KT 297, 2006).

Ogólna koncepcja schematu aplikacyjnego obejmuje trzy następujące etapy:

- zdefiniowanie zakresu tematycznego modelu oraz jego harmonizację z innymi modelami;
- sformalizowanie opisu modelu w postaci tzw. schematu semantycznego;
- integrację modelu danych ze schematami znormalizowanymi (geometrii i topologii, odniesień przestrzennych i czasowych, jakości i metadanych).

We wszystkich tych etapach główne znaczenie dla budowy schematu aplikacyjnego ma definicja obiektu (ang. *feature*), zwanego również wyróżnieniem (rozdz.2):

Obiekt jest to *abstrakcja zjawisk występujących w świecie rzeczywistym* (ISO, 2002; KT 297, 2006).

W dokumentach INSPIRE (INSPIRE, 2008a) oraz w ustawie o infrastrukturze informacji przestrzennej występuje pojęcie **obiektu przestrzennego** jako abstrakcyjnej reprezentacji przedmiotu, zjawiska fizycznego lub zdarzenia związanego z określonym miejscem lub obszarem geograficznym.

Obiektem jest zatem wyobrażenie lub reprezentacja dowolnego bytu w świecie rzeczywistym, przy czym może nim być w zależności od kontekstu zarówno *typ obiektu* jako reprezentacja pewnej kategorii lub grupy bytów o podobnych właściwościach, jak też *przypadek obiektu*, czyli pewien konkretny taki byt, o skonkretyzowanych właściwościach. Na przykład obiekt **budynek** może oznaczać zarówno grupę bytów opisywanych wspólnymi kategoriami cech, jak m.in. liczba kondygnacji (typ *obiektu budynek*), jak też, w innym kontekście, konkretny budynek o konkretnych wartościach tych cech, np. liczba kondygnacji równa 5 (*przypadek obiektu budynek*).

Każdy typ obiektu występujący w schemacie aplikacyjnym winien mieć swoją nazwę i definicję oraz być umieszczony w katalogu obiektów. Reguły dotyczące katalogowania i definiowania typów obiektów zawiera norma ISO (2005b).

Z pojęciem obiektu jest związanych wiele innych pojęć, jak *atrybut obiektu*, *asocjacja obiektu* oraz *operacja obiektu*, których definicje podane są poniżej.

Atrybut obiektu jest to *właściwość (cecha charakterystyczna) obiektu* (ISO, 2002; INSPIRE, 2008a; KT 297, 2006).

Atrybuty opisują wszelkie statyczne (niezmienne w czasie) cechy obiektu, jak na przykład jego identyfikator, jakość danych, położenie. Mogą one występować jako *typy atrybutów*, gdy dotyczą typów obiektów, albo jako *przypadki atrybutów*, gdy dotyczą przypadków

⁶ Jako *aplikację* należy przy tym rozumieć proces, w którym używa się danych i/lub przetwarza je zgodnie z oczekiwaniami użytkownika przy użyciu systemu komputerowego (KT 297, 2006).

obiektów. Atrybut występujący jako typ jest identyfikowany przez przypisaną mu nazwę oraz ma również przypisany typ danych, czyli dziedzinę wartości jakie może przechowywać. Atrybut występujący jako przypadek może posiadać także konkretną wartość z tej dziedziny. Na przykład, atrybut o nazwie **liczKond**, odpowiadający liczbie kondygnacji obiektu budynek, może mieć przypisany typ danych **Integer** jako dziedzinę wartości liczb całkowitych. Dla konkretnego budynku wartością tego atrybutu może być np. 5. Typy atrybutów obiektów powinny mieć definicje i być umieszczone w katalogu obiektów zgodnie z (ISO, 2005b).

Asocjacja obiektu jest to związek, który łączy przypadki pewnego typu obiektu z przypadkami tego samego lub innego typu obiektu (ISO, 2005a).

Asocjacja jest także nazywana *powiązaniem* obiektów. Umożliwia ona wzajemne kojarzenie przypadków różnych typów obiektów, co kolokwialnie można wyrazić jako to, „co jeden konkretny obiekt danego typu ma do zrobienia z jednym lub wieloma obiektami innego typu”. Podobnie jak obiekty i atrybuty, asocjacje winny mieć unikalne nazwy i definicje. Asocjacje są dwustronne, tj. od typu obiektu **A** do **B**, jak też od **B** do **A**. Przykładem może być konkretny budynek, który **stoiNa** jednej lub wielu działkach, jak też działka, która **zawiera** zero lub wiele budynków (wyróżniono potencjalne nazwy tej asocjacji). Podobnie, jak typy obiektów i atrybutów obiektów, również typy asocjacji powinny mieć definicje i być umieszczone w katalogu obiektów zgodnie z (ISO, 2005b).

Operacja obiektu jest to działanie, jakie może wykonać każdy obiekt danego typu (ISO, 2002; KT 297, 2006).

Schemat aplikacyjny zwiera typy obiektów, typy atrybutów obiektów, typy związków obiektów oraz typy ograniczeń obiektów. Opis ten jest często nazywany schematem semantycznym i powinien być powiązany (zintegrowany) z następującymi schematami znormalizowanymi, które specyfikują (rys. 1.13):

- *systemy odniesień* dla opisanego położenia i czasu według norm: PN-EN-ISO 19111 (odniesienia przestrzenne za pomocą współrzędnych), PN-EN-ISO 19112 (odniesienia przestrzenne za pomocą identyfikatorów geograficznych) oraz PN-EN-ISO 19108 (aspekt czasowy);
- *typy obiektów przestrzennych*, reprezentujących aspekty przestrzenne (geometrię i topologię) obiektów geograficznych, według normy PN-EN-ISO 19107;
- tzw. *elementy jakości danych* (wymierne) oraz *elementy przeglądowe jakości danych* (opisowe) zawierające informację o jakości konkretnych obiektów, atrybutów i związków, według norm PN-EN-ISO 19113 i PN-EN-ISO 19114;
- *metadane*, czyli dane o danych, charakteryzujące dany zbiór danych lub dowolny jego fragment, według PN-EN-ISO 19115.

Norma PN-EN-ISO 19109 zawiera reguły powiązania schematu aplikacyjnego ze schematami znormalizowanymi, których rygorystyczne przestrzeganie jest niezbędne dla zapewnienia możliwości zautomatyzowanego przetwarzania i wymiany danych, a które tym samym są jednym z koniecznych elementów współdziałania.

1.5. Schemat przestrzenny: opisywanie geometrii, topologii i położenia

Rysunek 1.14 przedstawia rolę schematu przestrzennego według PN-EN-ISO 19107 w opisywaniu położenia obiektu przestrzennego poprzez jego typ, przy czym klasa **Obiekt przestrzenny** reprezentuje tu przestrzenne (geometria, topologia i położenie za pomocą współrzędnych) aspekty obiektu geograficznego.

Klasa **System odniesienia** (rys. 1.15) jest kategorią ogólną, której specjalizacjami są klasy **System odniesień przestrzennych** oraz **System odniesień czasowych**. Z kolei, specjalizacją pierwszej z nich jest **System odniesień za pomocą współrzędnych**, którego schemat pojęciowy jest przedstawiony w normie ISO 19111 (ISO, 2007) i na rysunku 1.18 oraz **system odniesień za pomocą identyfikatorów geograficznych**, którego schemat pojęciowy jest przedstawiony w normie ISO 19112 (ISO, 2002). Schemat pojęciowy systemu odniesień czasowych jest przedstawiony w normie ISO 19108 (ISO, 2003).

Położenie obiektu geograficznego nazywamy:

- **położeniem bezpośrednim**, gdy jest zdefiniowane za pomocą współrzędnych w stosownym systemie odniesienia;
- **położeniem pośrednim**, gdy jest zdefiniowane za pomocą identyfikatorów geograficznych w stosownym systemie odniesienia, oraz
- **położeniem czasowym**, gdy jest zdefiniowane na osi czasu w stosownym systemie odniesienia.

Właściwości przestrzenne obiektów geograficznych, czyli ich geometria i topologia oraz położenie za pomocą współrzędnych, opisywane są przez powiązanie typu tego obiektu z odpowiednim typem obiektu przestrzennego – geometrycznym lub/i topologicznym. Na przykład obiekty geograficzne takie, jak studzienka, linia energetyczna i działka, mogą być powiązane odpowiednio z obiektami punkt, krzywa i powierzchnia w modelu geometrii lub/i z obiektami węzeł, krawędź i obszar w modelu topologii.

Kompletny zestaw modeli pojęciowych geometrii i topologii dla wektorowej reprezentacji danych geograficznych w przestrzeni 0-, 1-, 2- i 3-wymiarowej zawiera norma PN-EN-ISO 19107 *Spatial schema* (ISO, 2003). Charakterystyki przestrzenne obiektów dane są przez wartości atrybutów typu geometrycznego (*GM_Object*) lub typu topologicznego (*TP_Object*). Model geometrii zawiera mechanizmy opisu ilościowego, tj. wynikające ze współrzędnych, charakterystyk przestrzennych takich, jak wymiar, wielkość, kształt, orientacja i położenie obiektów. Aspekty geometryczne są jedynymi cechami informacji geograficznej, które ulegają zmianie wraz ze zmianą geodezyjnego układu odniesienia. Topologia, natomiast, daje przestrzenne charakterystyki jakościowe obiektów niezmiennie względem ciągłych zmian geodezyjnego układu odniesienia, takie, jak m.in. rozłączność, przyleganie, sąsiedztwo, przecinanie się czy zawieranie się jednych obiektów geograficznych w innych.

Główną konstrukcją geometryczną według (ISO, 2003) jest klasa *GM_Object*, która jest abstrakcyjną generalizacją ponad 50 klas – obiektów specjalizowanych. Dla celów większości zadań praktycznych, ograniczonych do opisu geometrii w przestrzeni co najwyżej dwuwymiarowej, sformułowano tzw. profil normy PN-EN-ISO 19107 jako normę PN-EN-ISO 19137 (ISO, 2004), w którym liczba podtypów klasy *GM_Object* wynosi sześć najczęściej używanych. Główne typy geometryczne według tej normy są przedstawione na diagramie (rys. 1.16). Wymienione na tym diagramie klasy pomocnicze:

`DirectPosition`, `GM_Position` i `GM_PointArray`, opisują położenie za pomocą współrzędnych (jako swych atrybutów) i są zdefiniowane w normie PN-EN-ISO 19107 (rys. 1.17).

Jak wynika z rysunku 1.17, opisywanie położenia za pomocą współrzędnych w każdym przypadku sprowadza się, poprzez ciąg kolejnych przywołań klas w roli typów danych, do użycia w tej roli także klasy `DirectPosition`, która ma bezpośrednie powiązanie z abstrakcyjną klasą `SC_CRS`⁷ schematu odwołań przestrzennych za pomocą współrzędnych w normie PN-EN-ISO 19111 (ISO, 2007). Schemat ten jest przedstawiony w uproszczonej postaci na rysunku 1.18.

Kluczowe znaczenie dla geometrycznej reprezentacji obiektów 0-wymiarowych (punktowych) ma klasa `GM_Point`, przedstawiona na rysunku 1.19. Wykorzystanie tej klasy dla opisu geometrii kilku przykładowych punktowych obiektów geograficznych zawartych w standardach technicznych Głównego Geodety Kraju ilustruje rysunek 1.20, gdzie klasa `OMG_ObjektPunktowy` jest abstrakcyjnym uogólnieniem geometrii tych obiektów, które mogą mieć odmienne definicje (co do list atrybutów i operacji) w poszczególnych standardach GJK. `OMG` jest tutaj hipotetycznym ogólnym modelem geodezyjnym, który stanowi abstrakcyjne uogólnienie odmiennych definicji typów obiektów geograficznych. Podobnie, klasa `OMG_Punkt` występująca w roli typu danych, jest specyficznym dla `OMG` odpowiednikiem `GM_Point`.

Podobne kluczowe znaczenie dla obiektów 1- i 2-wymiarowych ma klasa `GM_LineString`, której definicję przedstawia rysunek 1.21. Geometria standardowych według GJK jednowymiarowych obiektów przestrzennych jako linii łamanych (otwartych) może być opisana z użyciem klasy `GM_LineString` w sposób podobny, jak poprzednio, zilustrowany na rysunku 1.22. (Nazwa `LamanaUogolniona` pochodzi z instrukcji K-1; definicja klasy `GM_PointArray` podana zaś jest na rysunku 1.17).

W przypadku obiektu przestrzennego dwuwymiarowego opisem jego granicy jest „linia łamana uogólniona zamknięta” (cyt. z instrukcji K-1), odpowiadająca klasie `OMG_LamanaUogolnionaZamkniete` i zdefiniowana według ISO 19107 jako klasa `GM_Ring`. Klasa ta zawiera warunek, że ciąg punktów kontrolnych `GM_PointArray` stanowi zamknięty cykl. Definicja obiektu powierzchniowego miałaby więc postać, jak na rysunku 1.23. Należy przy tym zauważyć, że klasa `GM_Ring` dopuszcza m.in. również zdefiniowanie „zbioru poligonów z enklawami”, przewidzianego instrukcją K-1.

Przytoczone powyżej diagramy należy traktować jedynie jako przykłady wielu możliwych sposobów opisu geometrii i położenia obiektów geograficznych z wykorzystaniem norm PN-EN-ISO 19107 i PN-EN-ISO 19137. Podobny mechanizm może znaleźć zastosowanie także w przypadku opisu topologii tych obiektów i sprowadza się do zidentyfikowania i przywołania odpowiedniej klasy lub klas ze schematu przestrzennego.

Alternatywną względem współrzędnych metodą opisu położenia obiektów przestrzennych jest opis ich lokalizacji w stosunku do innych obiektów. Opis taki może przyjmować m.in. formę zawierania się jednego obiektu w innym (np. województwo – powiat – miejscowość – ulica – numer domu), bieżącej miary wzdłuż obiektu liniowego (np. 45 km drogi E7), bądź też luźnej informacji (np. restauracja leży pomiędzy muzeum a kinem). Tego typu formy opisu położenia są usystematyzowane w normie PN-EN-ISO 19112 (ISO, 2002) i noszą miano identyfikatorów geograficznych.

⁷ *Coordinate Reference System.*

Kluczowym dla takiego opisu jest system odniesienia za pomocą identyfikatorów geograficznych, który jest oparty na hierarchicznym podziale terytorium, jak np. „powiat – miasto – adres”. Z systemem tym, przedstawionym na rysunku 1.24, związane są m.in. następujące pojęcia:

- gazeter (**SI_Gazetteer**) – zbiór identyfikatorów wszystkich obiektów przestrzennych w powiązaniu z ich opisowymi lokalizacjami; lokalizacjom tym mogą być przypisane, w miarę potrzeby, ich położenia w układzie odniesienia za pomocą współrzędnych według ISO 19111;
- typ lokalizacji (**SI_LocationType**) – jednostka terytorialna w danym systemie odniesienia, jak np. obszar administracyjny, miasto, dzielnica, ulica i nieruchomość;
- instancja lokalizacji (**SI_LocationInstance**) – obiekt przestrzenny zarejestrowany w gazeterze, któremu jest przypisany jeden lub więcej typów lokalizacji.

Na rysunku 1.25 podany jest przykład opisu położenia działki z użyciem klasy **SI_LocationInstance** w roli typu danych atrybutów adresowych.

1.6. Integracja schematu aplikacyjnego ze schematami znormalizowanymi

Istota integracji schematu aplikacyjnego ze schematami standardowymi, zawartymi w normach, sprowadza się do wykorzystania w budowanym schemacie zawartych w normach schematów pojęciowych (lub ich fragmentów) dla typowych i często stosowanych zagadnień (jak np. geometria i topologia, jakość, położenie itp.). Istotę tę ilustruje rysunek 1.26, na którym za pomocą diagramu pakietów wyrażono wykorzystanie różnych schematów standardowych w budowanym schemacie aplikacyjnym.

Jako realizację zależności <<uses>> można wskazać co najmniej kilka metod pozwalających powiązać dany schemat aplikacyjny użytkownika z dowolnym innym schematem, w tym ze schematem znormalizowanym, przy czym przeważnie jest to powiązanie ukryte lub jawne, odpowiednich klas obu schematów. Do najprostszych spośród tych metod należą:

- przywołanie klasy ze schematu znormalizowanego w roli typu danych atrybutu w budowanym schemacie, zilustrowane na rysunku 1.27. Typy danych **GM_Object**, **EX_GeographicBoundingBox** i **MD_LegalConstraint** pochodzą odpowiednio z pakietu geometrii w ISO 19107 *Schemat przestrzenny* oraz z pakietów rozciągłości i metadanych w ISO 19115 *Metadane*.
- połączenie klasy w budowanym schemacie z odpowiednią klasą schematu standardowego za pomocą związku powiązania, agregacji, kompozycji lub zależności, jak na rysunku 1.28. Pełna konfiguracjaPrzestrzenna instancji klasy działka jest tu opisana za pomocą klasy **GM_Complex** w schemacie przestrzennym ISO 19107.
- wyspecyfikowanie klasy w budowanym schemacie jako specjalizacji klasy schematu standardowego z użyciem związku dziedziczenia (generalizacji), jak na rysunku 1.29. Nowa klasa podtypu dziedziczy wszystkie właściwości (atrybuty, operacje, ograniczenia i związki) klasy nadtypu, a ponadto umożliwia wyspecyfikowanie własnych właściwości. Jest to zatem integracja rozszerzająca zasób informacji klasy znormalizowanej. Przykład ten opisuje klasę **HistoriaBudyunku** w postaci ciągu zdarzeń, zdefiniowanych jako wyliczeniowy typ danych <<CodeList>> **Zdarzenie**. Oprócz tego klasa **HistoriaBudyunku** zdefiniowana jest jako specjalizacja klasy **TM_Topological**

Complex⁸, pochodzącej z pakietu (modelu) czasowego normy ISO 19108:2002 *Schemat czasowy*.

Norma PN-EN-ISO 19109:2005 podaje bardziej szczegółowe reguły integracji schematu aplikacyjnego użytkownika ze schematami metadanych, jakości danych, odniesień czasowych, geometrii i topologii oraz identyfikatorów geograficznych.

Kojarzenie ze sobą schematów aplikacyjnych dla różnych zakresów tematycznych informacji geograficznej wymaga częstokroć uzgodnienia list atrybutów dla odmiennie zdefiniowanych typów obiektów reprezentujących te same kategorie pojęciowe informacji geograficznej. Typowym przykładem mogą być tutaj instrukcje techniczne GGK (np. K-1, G-5, G-7, TBD), odmiennie definiujące takie pojęcia, jak budynek, działka, obręb, itp. Sytuację tę ilustruje przykładowo rysunek 1.30, który jednocześnie przedstawia metodę takiego uzgodnienia (harmonizacji) poprzez zdefiniowanie odpowiedniej klasy nadtypu i wykorzystanie mechanizmu dziedziczenia.

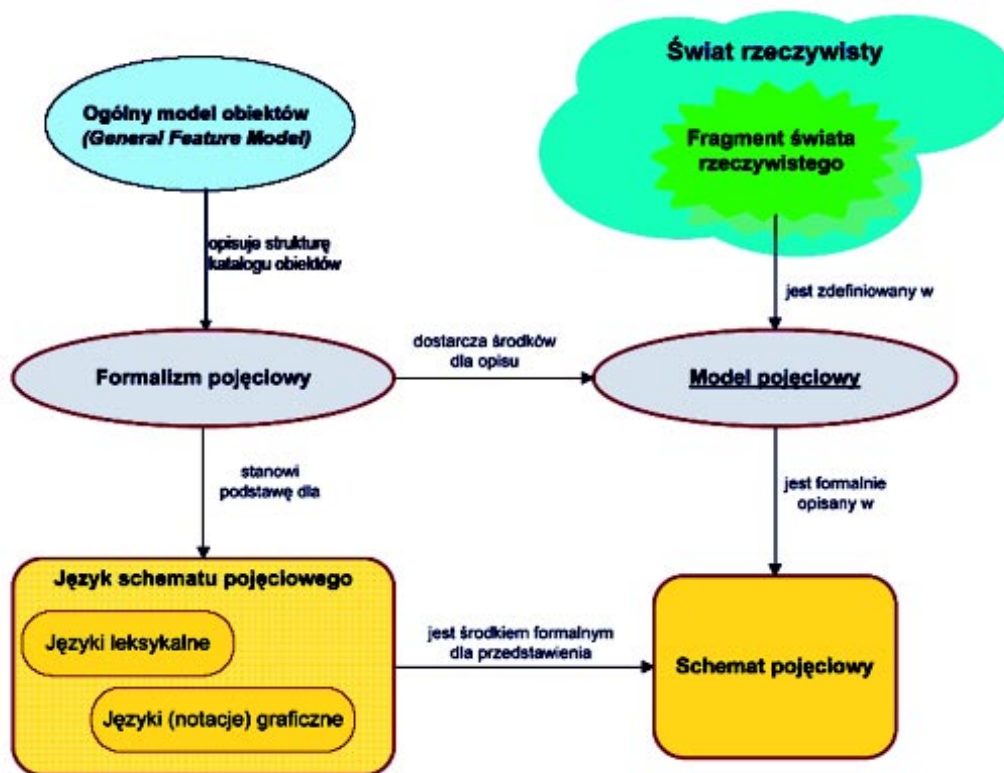
1.7. Przykład modelu pojęciowego

Na rysunku 1.31 przytoczony jest fragment modelu GESUT według instrukcji technicznej G-7 wydanej przez Głównego Geodetę Kraju jako przykład schematu aplikacyjnego według normy PN-EN-ISO 19109.

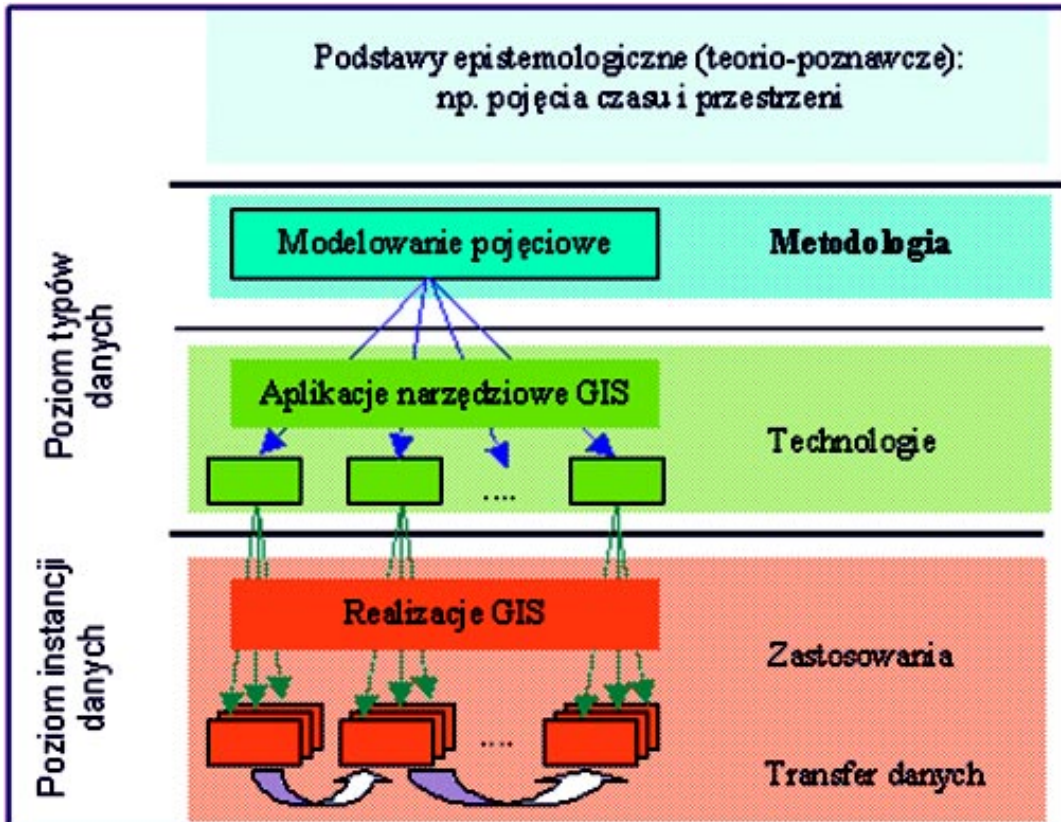
Literatura

- Booch G, Rumbaugh J, Jacobson I., 2002: UML przewodnik użytkownika, WNT, Warszawa.
- INSPIRE, 2007: *Dyrektywa 2007/2/WE Parlamentu Europejskiego i Rady z dnia 14 marca 2007 r. ustanawiająca infrastrukturę informacji przestrzennej we Wspólnocie Europejskiej (INSPIRE)*. Dziennik Urzędowy Unii Europejskiej, L 108, także:
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:PL:PDF>
- INSPIRE, 2008: Glossary. INSPIRE, <http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY>.
- ISO, 2002: ISO 19101:2002(E): Geographic information — Reference model; także jako: PN-EN-ISO 19101:2005, Informacja geograficzna – Model odniesienia.
- ISO, 2003: ISO 19112:2003: Geographic information – Spatial referencing by geographic identifiers
- ISO, 2004: ISO/DIS 19137: Geographic information – Generally used profiles of the spatial schema and of similar important other schemas.
- ISO, 2005a: ISO 19109:2005: Geographic information – Rules for application schema; także jako: PN-EN-ISO 19109:2009, Informacja geograficzna – Reguły schematów aplikacyjnych.
- ISO, 2005b: ISO 19110:2005: Geographic information – Methodology for feature cataloguing; także jako: PN-EN-ISO 19110:2006, Informacja geograficzna — Metodyka katalogowania obiektów.
- ISO, 2005c: ISO/TS 19103:2005: Geographic information – Conceptual Schema Language.
- ISO, 2007: ISO 19111:2007: Geographic information – Spatial referencing by coordinates.
- KT 297, 2006: e-Przewodnik do Polskich Norm w dziedzinie informacji geograficznej. Zespół PKN/KT 297, <http://www.e-przewodnik.gugik.gov.pl/>
- Pachelski W., Parzyński Z., 2007: Aspekty metodyczne wykorzystania norm serii ISO 19100 do budowy geodezyjnych składników krajowej infrastruktury danych przestrzennych. Konf. PTIP na temat Współpraca i koordynacja w zakresie geoinformacji dla zrównoważonego rozwoju w Polsce i Europie, Warszawa, 6-8 listopada 2007, oraz *Roczniki Geomatyki*, t. V, z. 3, 2007, ss. 113-121. PTIP, Warszawa.
- Schenck D., Wilson P. 1994: Information Modelling: the EXPRESS Way. Oxford Univ. Press.

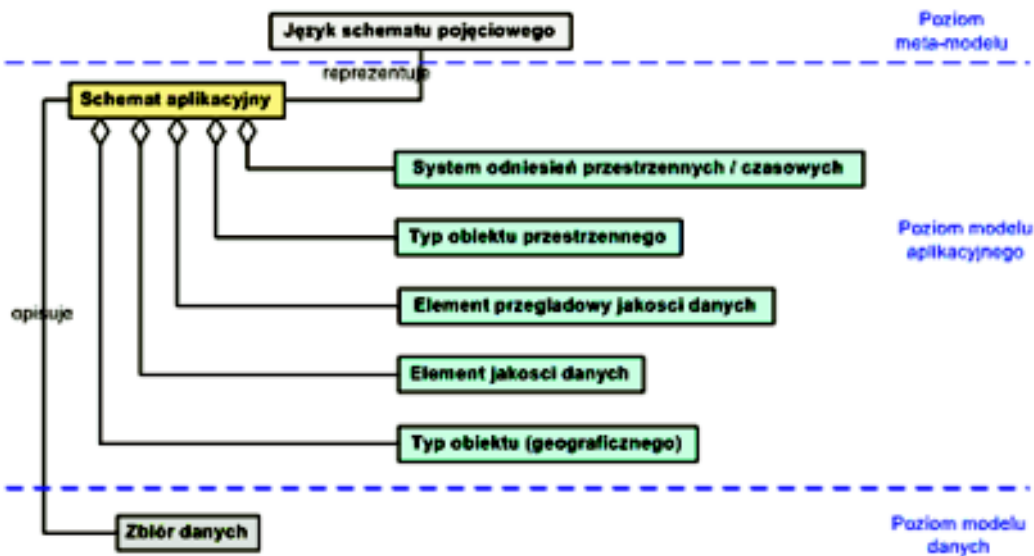
⁸ Złożona konstrukcja topologiczna (w schemacie czasowym).



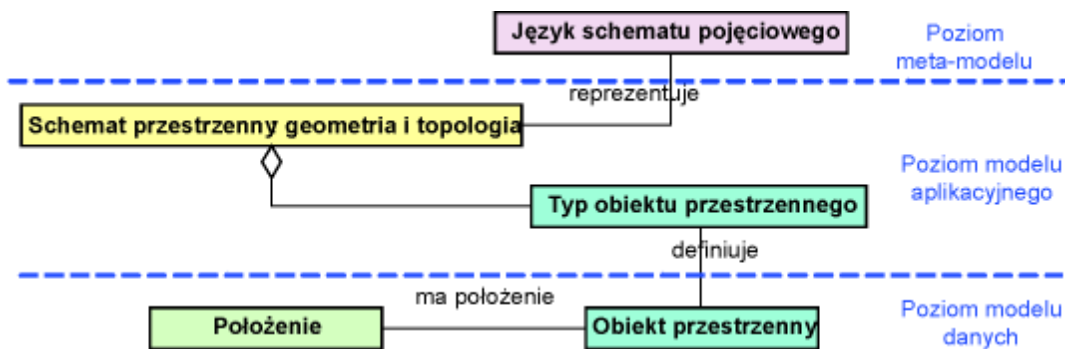
Rys. 1.3. Modelowanie pojęciowe informacji geograficznej (na podstawie: INSPIRE, 2008/D2.5)



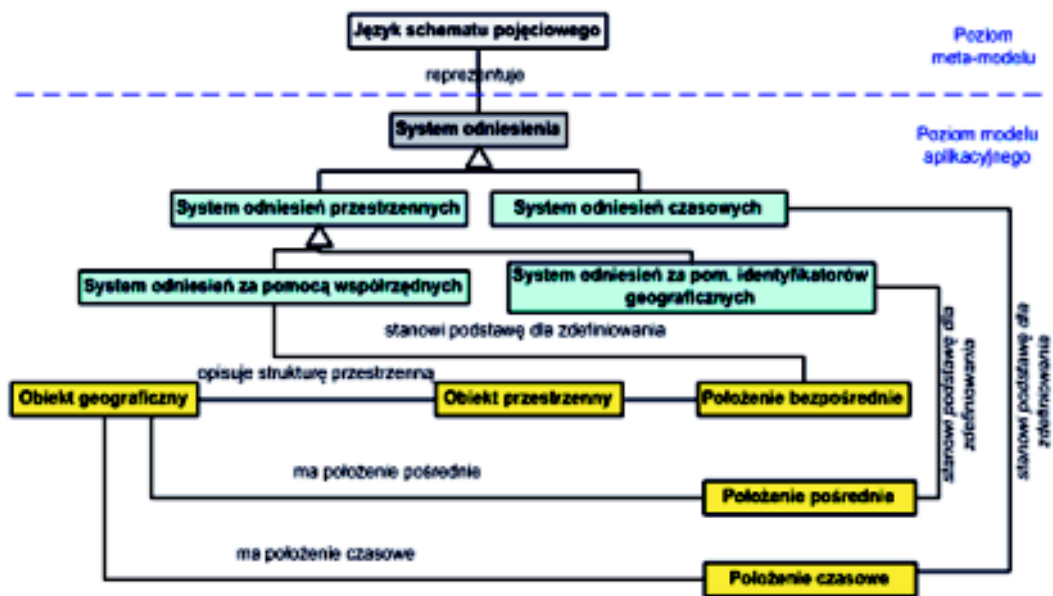
Rys. 1.4. Metodologia, technologie i zastosowania GIS



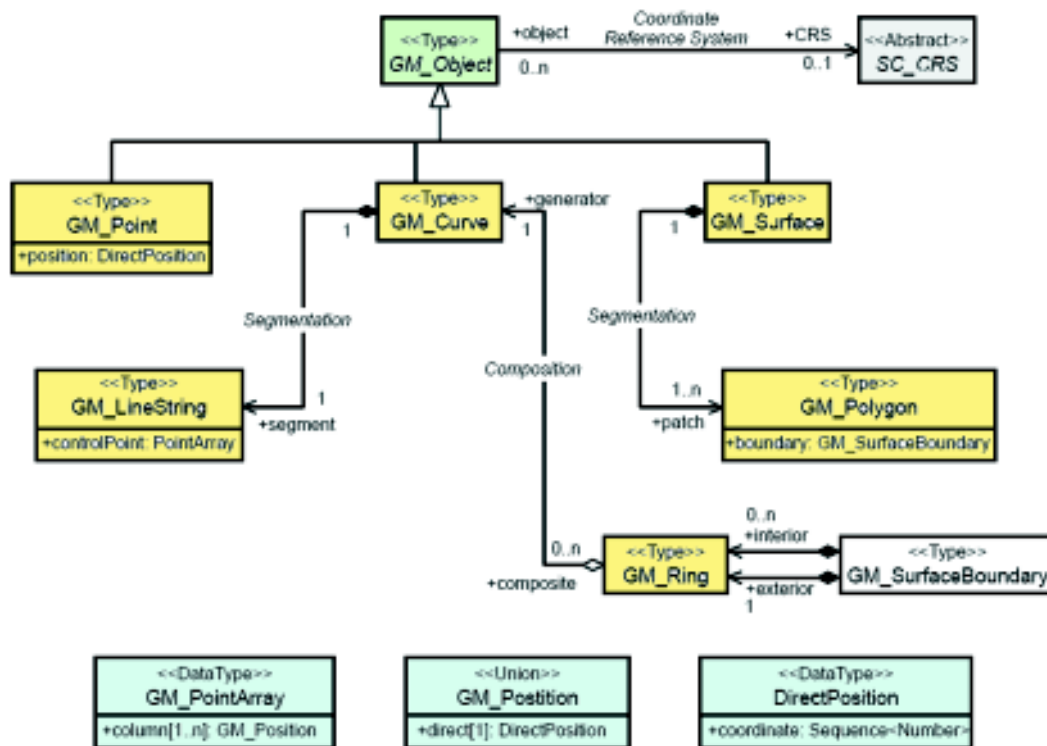
Rys. 1.13. Struktura schematu aplikacyjnego (wg ISO, 2002)



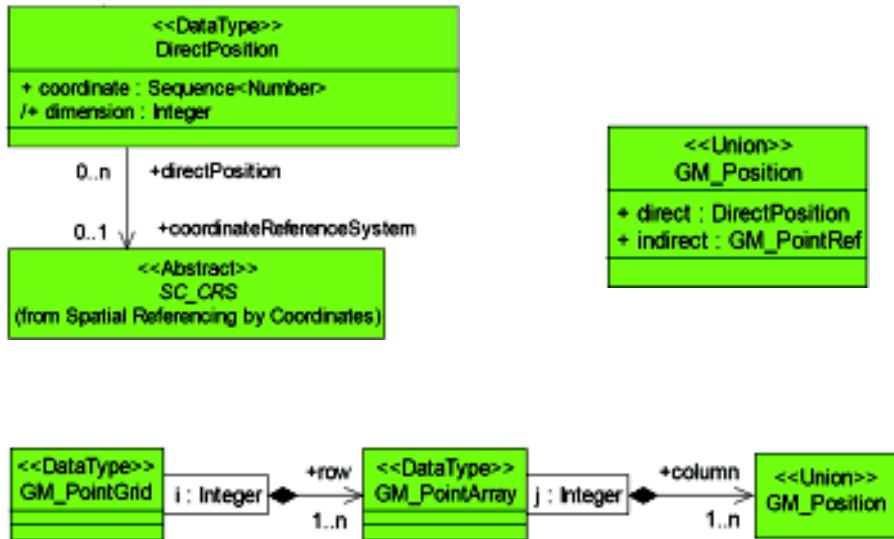
Rys. 1.14. Struktura opisu położenia obiektu przestrzennego (wg ISO, 2002)



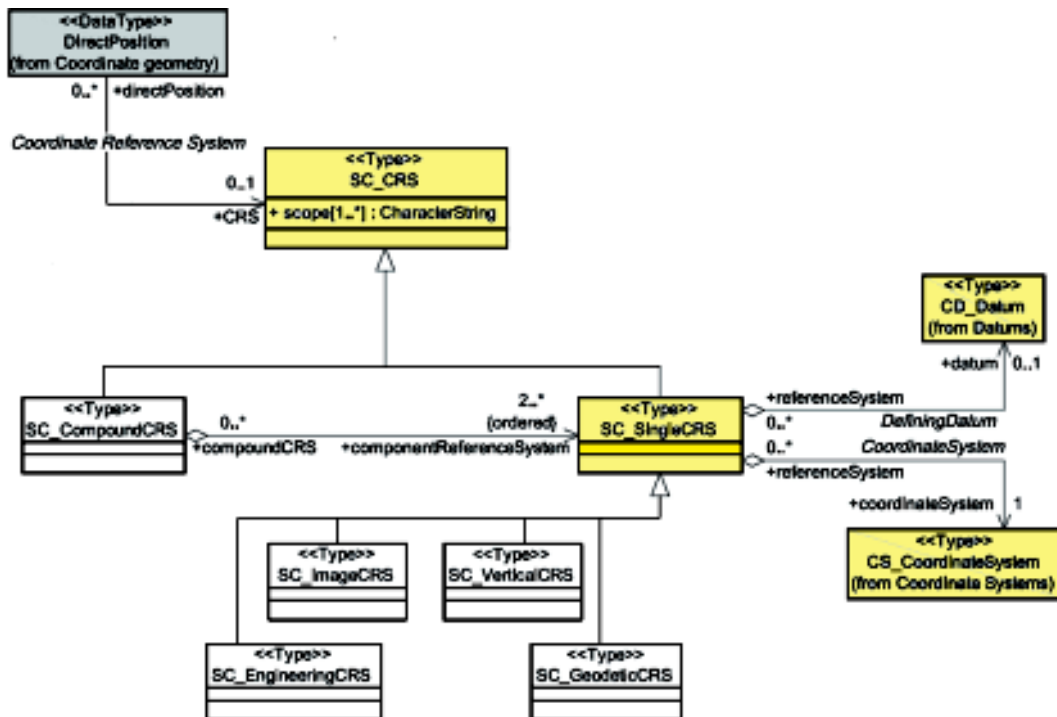
Rys. 1.15. Systemy odniesienia i opisywanie położenia (źródło: ISO, 2002)



Rys. 1.16. Główne struktury geometryczne (źródło: ISO, 2004)

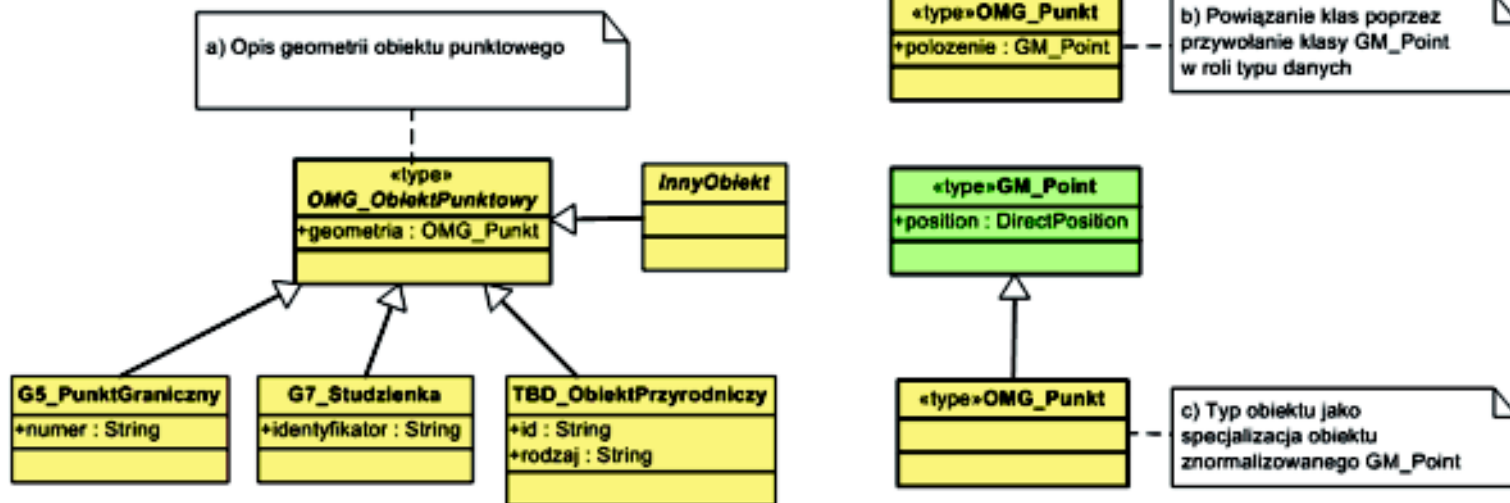
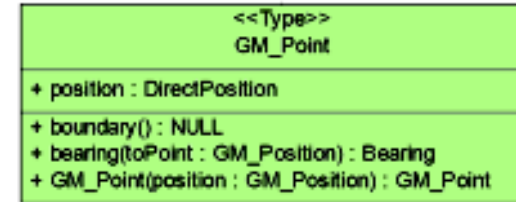


Rys. 1.17. Definicje pomocniczych klas dla opisywania położenia (źródło: ISO, 2003)



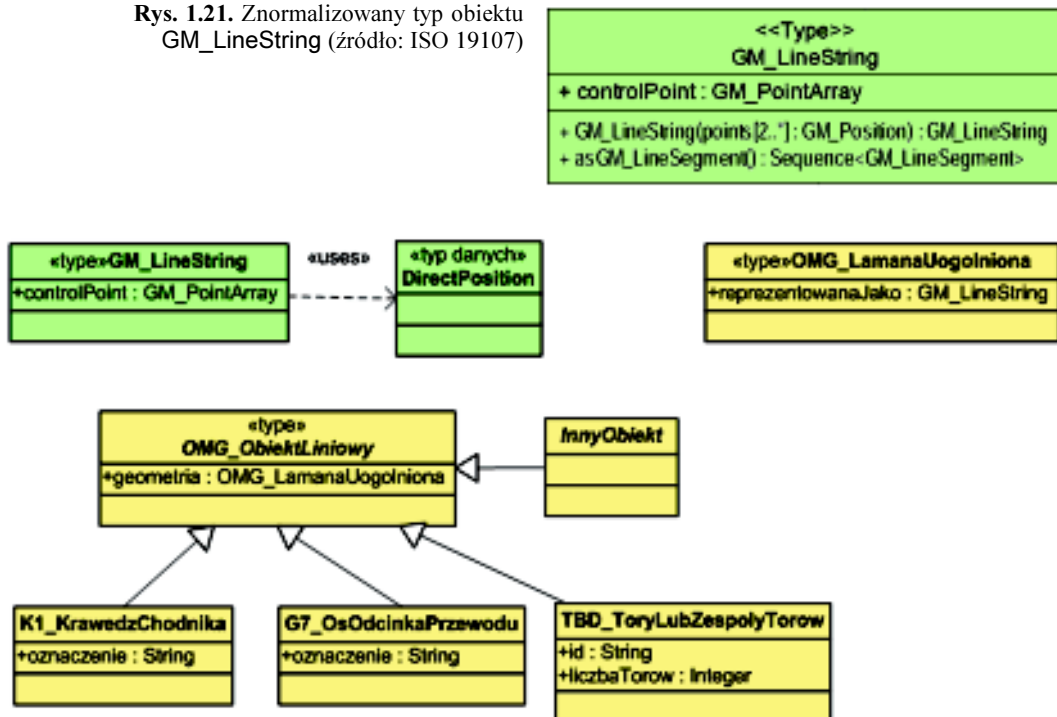
Rys. 1.18. System odniesień przestrzennych za pomocą współrzędnych (źródło: ISO 19111)

Rys. 1.19. Znormalizowany typ obiektu GM_Point (źródło: ISO 19107)

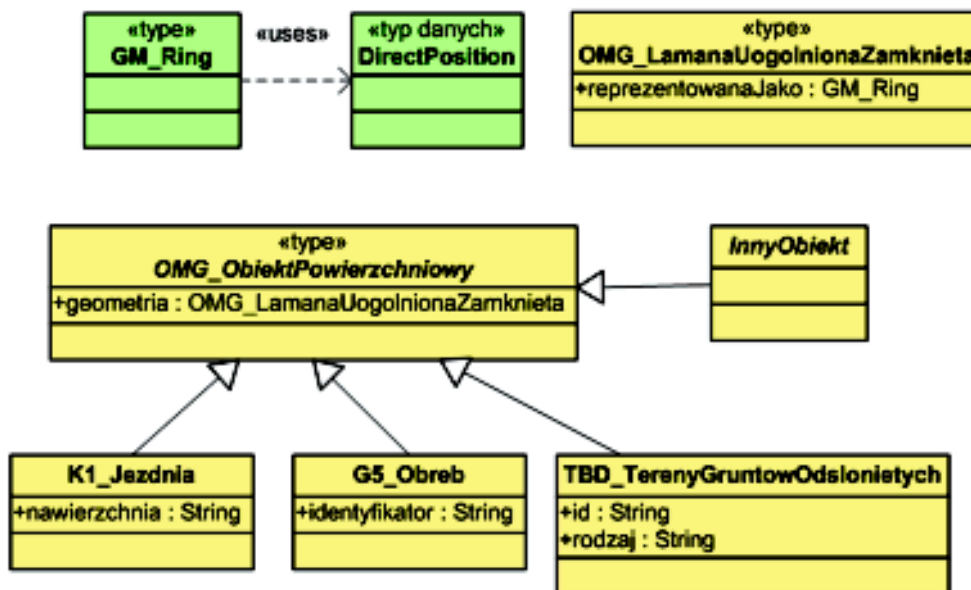


Rys. 1.20. Opis geometrii obiektu punkowego (źródło: Pachelski i Parzyński, 2007)

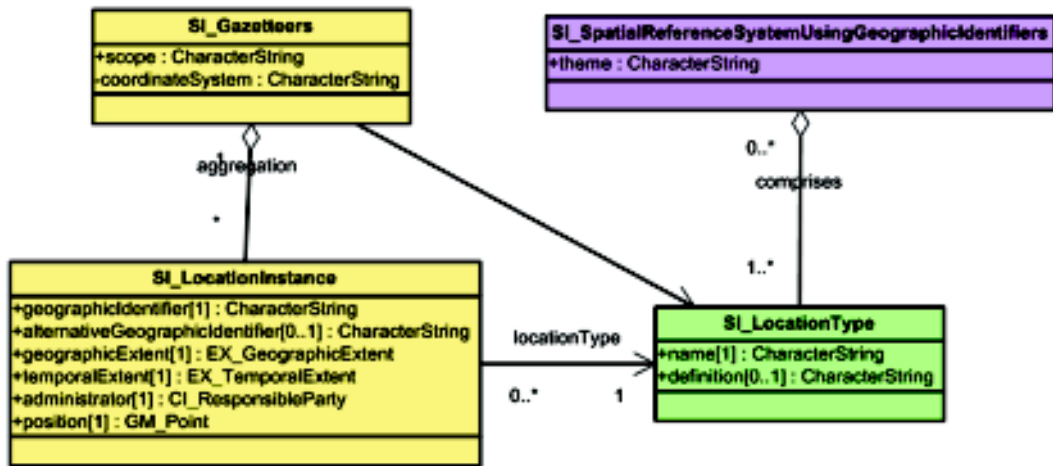
Rys. 1.21. Znormalizowany typ obiektu GM_LineString (źródło: ISO 19107)



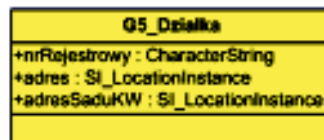
Rys. 1.22. Opis geometrii obiektu liniowego (źródło: Pachelski i Parzyński, 2007)



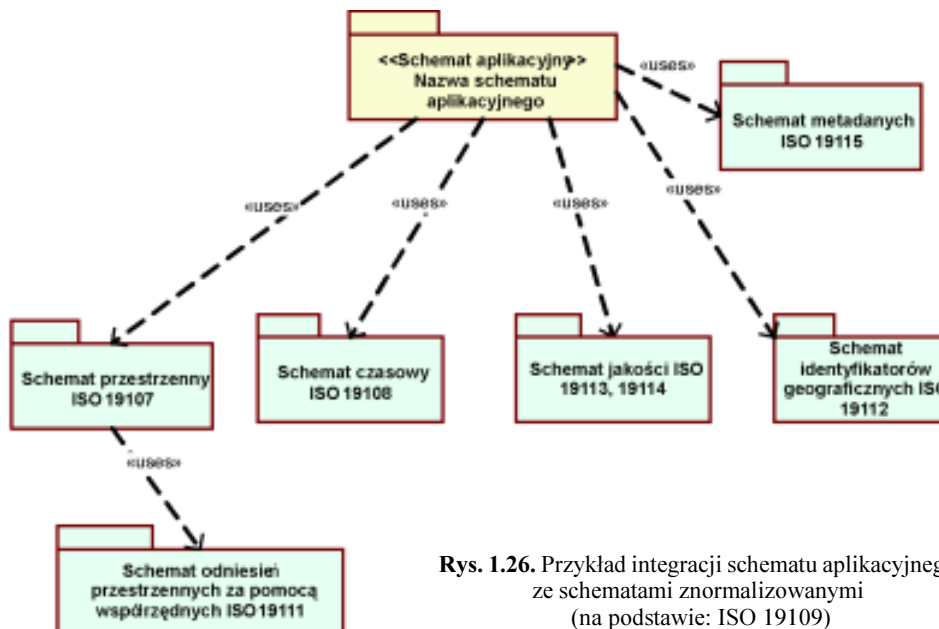
Rys. 1.23. Opis geometrii obiektu powierzchniowego (źródło: Pachelski i Parzyński, 2007)



Rys.1. 24. Schemat struktury odwołań z użyciem identyfikatorów geograficznych (wg ISO 19112)



Rys.1. 25. Przykład odwołania do identyfikatorów geograficznych

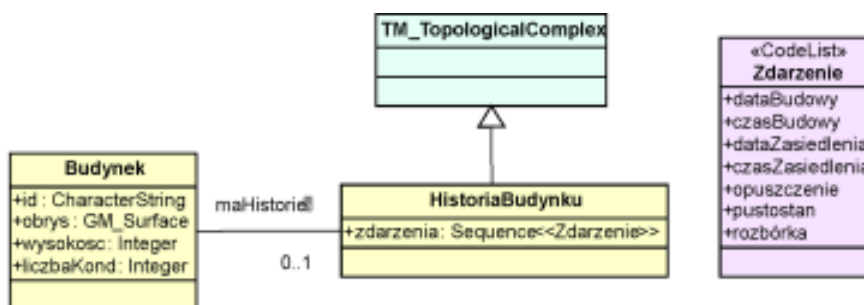


Rys. 1.26. Przykład integracji schematu aplikacyjnego ze schematami znormalizowanymi (na podstawie: ISO 19109)

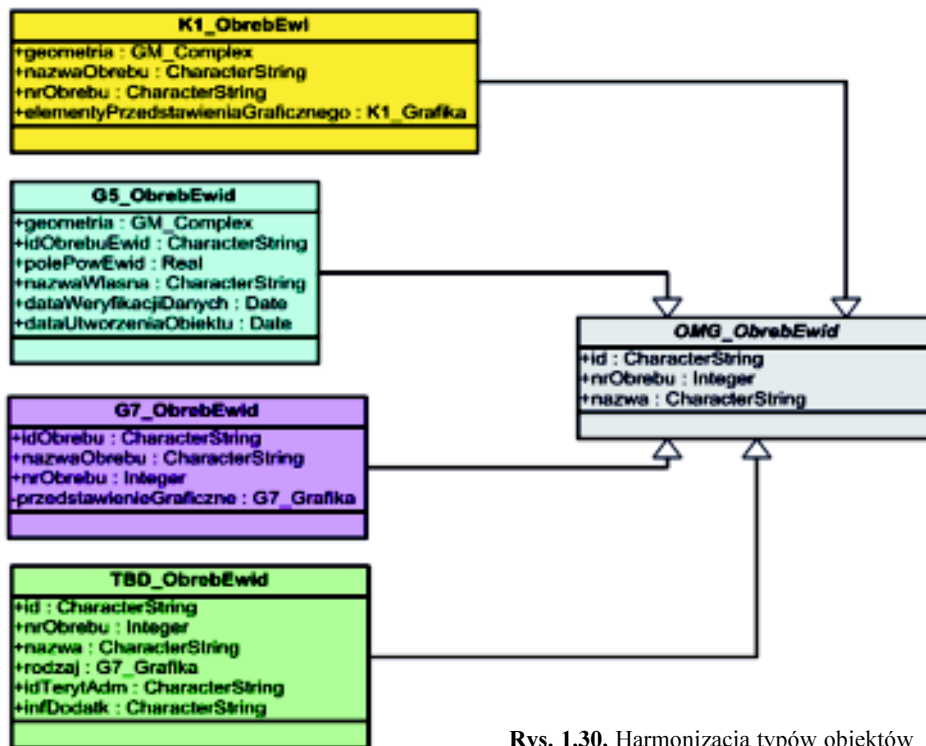
Rys. 1.27. Przykład przywołania klas ze schematów standardowych w roli typów danych atrybutów (na podstawie: ISO 19109)



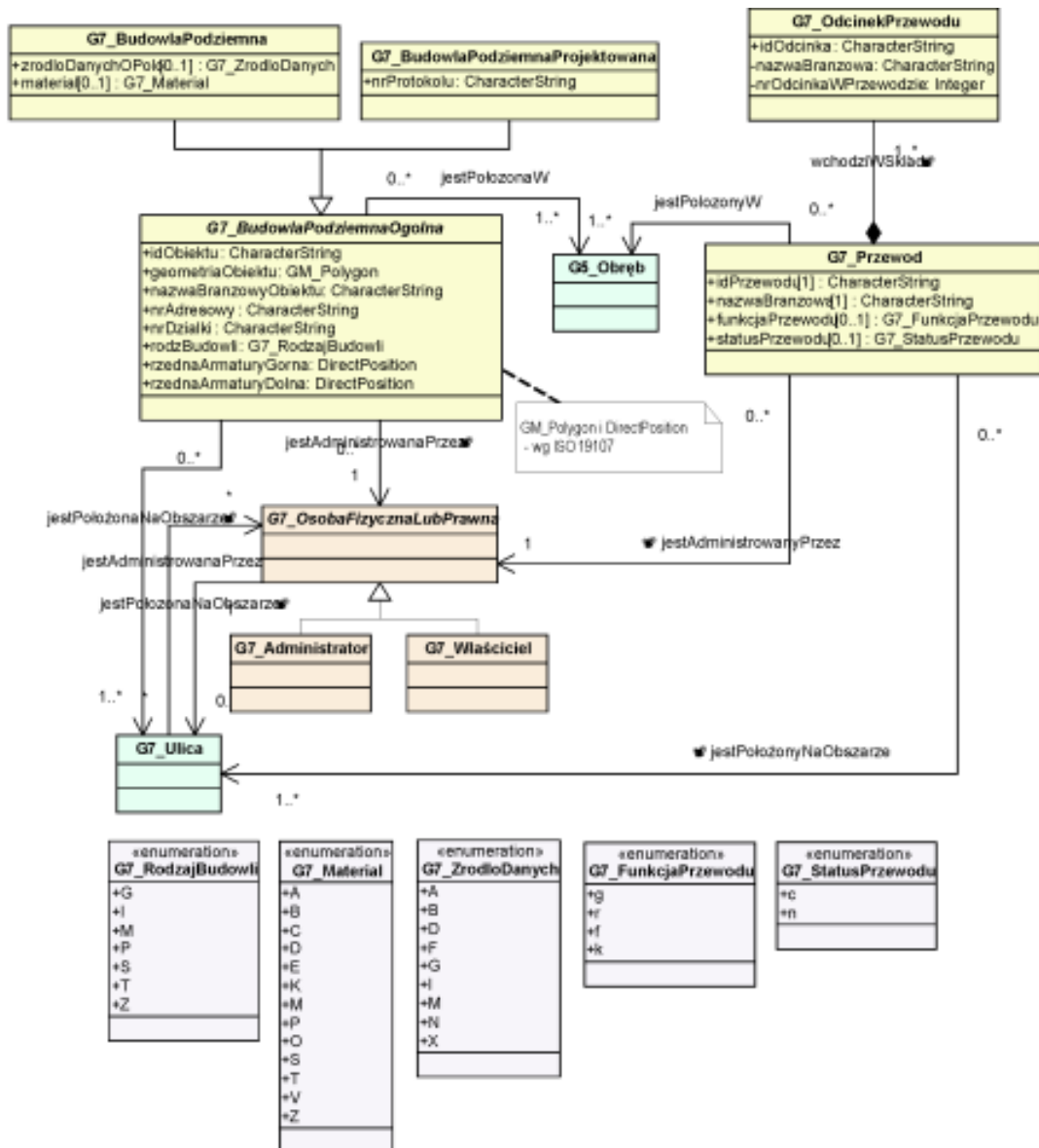
Rys. 1.28. Przykład powiązania klasy w schemacie budowanym z klasą schematu standardowego



Rys. 1.29. Przykład specyfikacji klasy jako specjalizacji klasy standardowej (na podstawie: ISO 19109)



Rys. 1.30. Harmonizacja typów obiektów



Rys. 1.31. Fragment modelu GESUT jako przykład schematu aplikacyjnego (źródło: Pachelski i Parzyński, 2007)

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