



**POLSKIE
TOWARZYSTWO
INFORMACJI
PRZESTRZENNEJ**

ROCZNIKI 2010

GEOMATYKI

Modelowanie danych przestrzennych

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

Janusz Michalak

2. TRANSFORMACJA POLSKICH DANYCH PRZESTRZENNYCH DO MODELI INSPIRE

Problematyka transformacji danych nie jest zagadnieniem nowym – pojawiła się już w początkowej fazie rozwoju systemów GIS. Zapisywanie danych przestrzennych, ze względu na ich złożone struktury, jest zadaniem trudnym i początkowo dominowały metody zapisu ściśle powiązane z konkretnymi systemami GIS, a przesyłanie danych pomiędzy różnymi systemami sprawiało wiele problemów. Najbardziej typowym sposobem rozwiązywania tych problemów były funkcje importu i eksportu wbudowywane w poszczególne systemy. Na tym etapie pojęcie modelu danych nie było znane i konsekwencją tego była najczęściej utrata części atrybutów, wynikająca z różnic pomiędzy sztywnymi strukturami plików danych stosowanych w różnych systemach GIS.

Jednym ze sposobów rozwiązania tych trudności było opracowywanie różnych, niezależnych od konkretnych programów GIS, sposobów zapisu danych dedykowanych albo określonym dziedzinom zastosowań albo określonym zasięgom terytorialnym. Przykładami tych drugich były narodowe formaty kodowania danych. W stosunkowo krótkim okresie powstały dziesiątki takich formatów, co wcale nie przyczyniło się do poprawy sytuacji.

Obecnie zagadnienia zapisu różnorodnych danych przestrzennych i ich transformacji pomiędzy różnymi modelami stanowią oddzielny dział geomatyki, jednak problemy jakie tam występują nie są wcale mniejsze niż te, na które napotymano dawniej. Główne zagadnienia jakimi zajmuje się ten dział geomatyki przedstawiono poniżej.

Modele struktur danych. Są różne dla różnych rodzajów zapisu geoinformacji, w tym w postaci:

- wyróżnień (*feature*), do których zalicza się szeroko pojęte obiekty geograficzne – najczęściej zapisywane wektorowo;
- pokryć (*coverage*) dla zapisu obrazów różnego typu przy użyciu danych rastrowych (*raster data*) lub danych macierzowych (*matrix data*) opisujących zmienność przestrzenną wielkości geometrycznych lub fizycznych, najczęściej zapisywanych w formie tablic tekstowych lub binarnych, a także w postaci wektorowej.

Gromadzenie danych przestrzennych w bazach danych. Ta problematyka w zastosowaniach do geoinformacji jest szczególnie trudna ze względu na znaczną różnorodność danych i złożoność ich struktur.

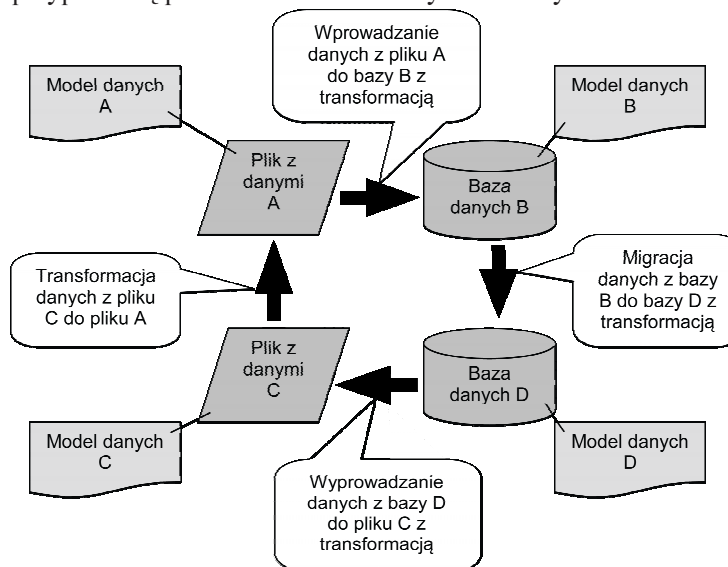
Zapis danych w plikach w sposób niezależny od oprogramowania i platform sprzętowo-systemowych, na których to oprogramowanie funkcjonuje. Dotyczy to, zarówno przechowywania danych jak i przesyłania ich pomiędzy różnymi systemami geoinformacyjnymi. W tym przypadku wielkie nadzieje są pokładane w językach znacznikowych bazujących na XML (*Extensible Markup Language*) (Bray, Paoli, Sperberg-McQueen, 1998; Mercer, 2000; Carlson, 2001). W zastosowaniach do danych przestrzennych językiem tym jest GML

(*Geography Markup Language*) (Lake, 1999). Zastosowanie tego języka do określonych danych dziedzinowych wymaga opracowania schematów aplikacyjnych XSD (*XML Schema Definition*), ponieważ jest on również jedynie bazą pozwalającą na opracowywanie różnych jego aplikacji, jakim są m.in. schematy dla poszczególnych tematów INSPIRE.

Transformacja danych pomiędzy różnymi formami ich zapisu – różniącymi się modelami struktur danych, układami odniesienia przestrzennego i sposobami kodowania danych liczbowych i tekstowych. Z transformacją danych mamy do czynienia m.in. gdy:

- wprowadza się dane z pliku do bazy danych i z reguły struktura bazy jest inna niż struktura pliku źródłowego;
- dokonywana jest migracja danych z jednej bazy do drugiej o innej strukturze;
- dane są wyprowadzane z bazy do pliku w celu przesłania ich do innego systemu geoinformatycznego, w którym będą analizowane lub przetwarzane;
- tworzy się nowy plik danych z jednego lub kilku plików źródłowych i w każdym przypadku mamy inną strukturę danych lub inną formę ich zapisu.

Te cztery przypadki są przedstawione schematycznie na rysunku 2.1.



Rys. 2.1. Cztery przypadki operowania na danych przestrzennych z udziałem transformacji pomiędzy różnymi modelami

Można sobie wyobrazić następujący eksperyment. Dane z pliku A o innym modelu danych niż struktura bazy B są wprowadzane do tej bazy z transformacją z modelu A do modelu B. Następnie odbywa się migracja danych z bazy B do bazy D, również z transformacją z modelu B do D. Kolejny etap to wyprowadzenie danych z bazy D do pliku C, któremu również towarzyszy transformacja. Ostatnia operacja to utworzenie nowego pliku A¹ przez transformację danych zawartych w pliku C do modelu A. Jest bardzo mało prawdopodobne, że zawartość nowego pliku A¹ będzie identyczna z tymi danymi, które plik A zawierał początkowo. Zagadnienie to należy do problematyki odwracalności procesu przetwarzania danych podczas ich transformacji. Często tylko transformacja z jednego modelu do drugiego i następnie odwrotna (z drugiego do pierwszego) wykazuje nieodwracalność tych procesów.

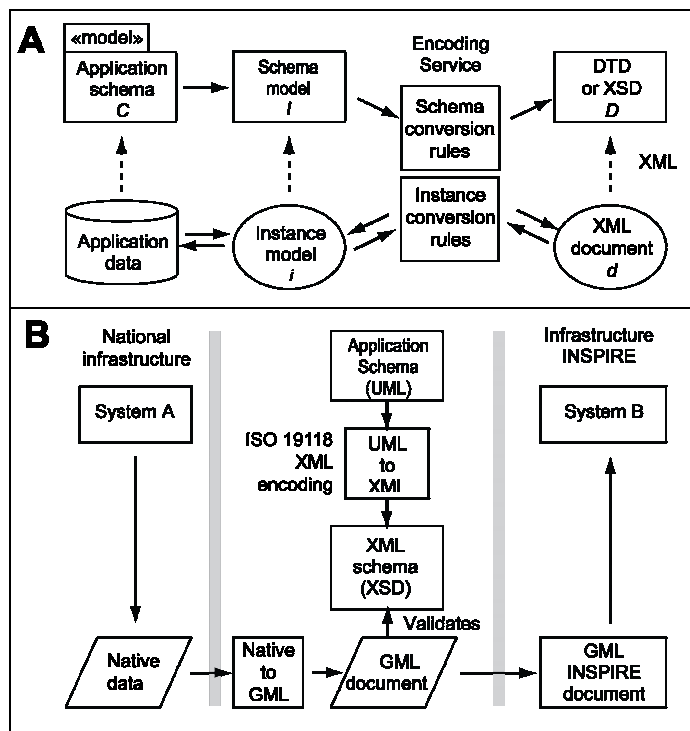
Powyższy przykład to tylko jeden z wielu problemów transformacji danych przestrzennych. W dalszej części tego rozdziału będzie przedstawionych wiele innych trudnych zagadnień dotyczących transformacji pomiędzy różnymi modelami danych:

- podstawowych różnic w zakresie technologicznym pomiędzy strukturami i formami danych polskich i danych zgodnych ze specyfikacjami INSPIRE;
- uwarunkowań procesu transformacji wynikających z regulacji prawnych określających implementację dyrektywy INSPIRE;
- zastosowanie języków XML i GML jako zapisu pośredniego w transformacji;
- transformacje zaawansowane zapisów XML i GML do form docelowych z zastosowaniem XSLT (*Extensible Stylesheet Language Transformations*).

2.1. Podstawowe pojęcia

Przekształcanie danych przestrzennych pomiędzy różnymi ich modelami nie jest zagadnieniem nowym. W pracach z lat dziewięćdziesiątych rozwijano tę problematykę. Na szczególną uwagę zasługują prace Skogana (1999 i inne we współautorstwie), które stanowią podstawę pojęciową metod transformacji danych przestrzennych. Do podstawowych pojęć z tego zakresu należą (rys. 2.2.):

- dane aplikacyjne (*Application data*) – dane odnoszące się do konkretnej aplikacji, w przypadku INSPIRE są to dane dotyczące jednego z tematów;
- schemat aplikacyjny (*Application schema*) – model UML opisujący semantykę, zawartość i logiczną strukturę danych aplikacyjnych;
- model schematu (*Schema model*) – podzbiór metamodelu UML przeznaczony do określenia schematu aplikacyjnego;
- model instancji (danych aplikacyjnych) (*Instance model*) – model będący profilem (podzbiorem) modelu schematu i reprezentujący konkretne dane aplikacyjne o strukturze logicznej odpowiadającej wymaganiom języka XML;
- dokument XML (*XML document*) – zapis danych zgodny z definicją semantyki, logicznej struktury wynikającej z modelu schematu po przekodowaniu do wymagań języka XML;
- usługa przekodowania (transformacji) (*Encoding Service*) – zbiór operacji wykonywanych przez serwer w celu zmiany formy i struktury logicznej danych aplikacyjnych do postaci docelowej w formie dokumentu (zbioru) XML – w przypadku danych przestrzennych jest to GML lub jego dziedzinowa specjalizacja;
- definicja typu dokumenty (*Document Type Definition – DTD*) lub ostatnio najczęściej – definicja schematu XML (XSD) – dokument określający semantykę i logiczną strukturę zapisu w dokumencie (pliku) XML;
- reguły konwersji schematu (*Schema conversion rules*) – zbiór reguł określających sposoby przekodowania modelu schematu na dokument definiujący strukturę zapisu XML (GML lub jego aplikacji) – dawniej był to zapis DTD, a obecnie stosuje się zapis w formie zgodnej z wymaganiami XSD;
- reguły konwersji instancji (danych aplikacyjnych) (*Instance conversion rules*) – zbiór reguł określających sposoby przekodowania samych danych – instancję (egzemplarz) danych, jako realizację modelu instancji, do dokumentu (zbioru) XML zgodnego z regułami kodowania określonymi w DTD lub XSD.



Rys. 2.2. Koncepcja transformacji danych pomiędzy różnymi modelami przedstawiona schematycznie: A – transformacja jest oparta na podziale modeli na dwa poziomy: model schematu (*schema model*) – bardziej ogólny model pojęciowy w UML i model instancji (*instance model*) – bardziej szczegółowy model ograniczony do wymagań specyfikacji języka XML (cytat z publikacji: Skogan, 1999), B – schemat transformacji danych pomiędzy dwoma systemami o różnych modelach danych (cytat z raportu Grupy Nordyckiej dostosowany do problematyki rozdziału)

Obecnie do tego podstawowego zestawu pojęć dochodzą nowe, związane z zastosowaniem nowych technologii w zakresie transformacji danych przestrzennych. Najważniejsze pojęcia przedstawiono poniżej.

Transformacja danych przestrzennych. Zmiana struktury logicznej i formy danych, często także zmiana treści przez pominięcie elementów danych lub uzupełnienie danymi z innych zbiorów. Można wyodrębnić kilka kategorii transformacji:

- transformacja współrzędnych z jednego układu odniesienia do innego, np. z EPSG:4326 (WGS84/Lat-Long) na EPSG:2180 (Poland CS 92);
- transformacja formatu zapisu liczb, przykładowo dla współrzędnych geograficznych z DMS (np. $52^{\circ}30'00''$) na DDD (np. 52.5000) lub zamiana przecinka dziesiętnego na kropkę dziesiętną;
- transformacja kodowania tekstów, np. z kodu ISO-8859-2 na UTF-8 (ośmiobitowy Unicode);
- transformacja jednostek wielkości geometrycznych i fizycznych, np. zmiana kilometrów na metry lub przeliczenie prędkości z km/h na m/s ;
- transformacja struktury danych z formatów innych niż aplikacje XML na zapisy oparte na XML;
- transformacja struktur danych pomiędzy różnymi aplikacjami XML, a w tym aplikacjami GML, także z wykorzystaniem technologii XSLT (Lehto, Sarjakoski, 2004; OSM, 2009; Galdos, 2002);
- szczególnym przypadkiem transformacji danych przestrzennych jest ich zamiana na zapis graficzny (w układzie arkusza papieru lub ekranu) za pomocą SVG (aplikacji

XML dla zapisu obrazów wektorowych) w celu zobrazowania tych danych – w tym przypadku również wykorzystuje się technologię XSLT (Tennakoon, 2003).

Transformacja danych przestrzennych może być realizowana w trzech kontekstach:

- jako transformacja *off-line* („na miejscu”) – operacja realizowana okresowo dla całego zbioru danych lub bazy danych i powstaje w jej wyniku nowy zbiór lub nowa baza danych;
- jako transformacja *on-line* („w locie”) – operacja transformacji jest jednym z ogniw łańcucha usług utworzonego przez usługę wywoływania innych usług; taki przypadek może wystąpić, gdy klient (system) wymaga zmiany struktury lub formy danych, ponieważ dane udostępniane nie spełniają jego potrzeb; transformacji poddana jest tylko ta część źródłowego zbioru danych, która jest przesyłana do klienta;
- jako usługa przetwarzania danych przestrzennych z zastosowaniem standardu WPS (*Web Processing Service*) (Michalak, 2010); w takim przypadku klient przesyła do serwera zbiór danych źródłowych lub wskazuje jego URL (*Uniform Resource Locator*) i określa lub wybiera reguły transformacji, natomiast serwer WPS dokonuje transformacji i przekazuje zbiór wynikowy lub podaje klientowi jego położenie za pomocą URL.

Przestrzeń nazw XML. Zakres logiczny funkcjonowania nazw elementów XML, zdefiniowanych w dokumencie lub kolekcji dokumentów XSD (dawniej DTD). Często dokumenty te są ogólniej nazywane słownikami XML, ponieważ zawierają określenia nazw występujących tam elementów i definicje struktur tych elementów, co w pełni odpowiada regułom słowników. Każdemu słownikowi przypisana jest przestrzeń nazw określona niepowtarzalnym prefiksem, co pozwala na używanie tych samych nazw w różnych słownikach stosowanych w określonej aplikacji. Na przykład element o nazwie „identyfikator” może wystąpić jednocześnie w więcej niż w jednym słowniku XML (schemacie aplikacyjnym) i w każdym słowniku może mieć inne znaczenie. Element „gml.identifier” może być nazwą innego elementu o innej semantyce i strukturze niż element o nazwie „geo.identifier” ponieważ pierwszy należy do przestrzeni nazw „gml” a drugi do przestrzeni „geo”. Mechanizm ten bardzo ułatwia budowanie schematów aplikacyjnych wykorzystujących inne już istniejące schematy, bo nie ma problemu z powtarzającymi się nazwami elementów. Przykładem takiej sytuacji jest język GWML (*Groundwater Markup Language*) wykorzystujący schematy GML i GeoSciML (*Geosciences Markup Language*), który też jest aplikacją języka GML.

Harmonizacja modeli danych (nazywana często dla uproszczenia „harmonizacją danych”). Procedura, której celem jest opracowanie jednego spójnego modelu struktury logicznej i semantyki dla wielu zbiorów danych, które zostały utworzone w różnych standardach dla różnych platform sprzętowo-systemowych i z tego powodu nie spełniają wymagań interoperacyjności. W konsekwencji tego, różne zbiory danych o różnych modelach mogą być przetransformowane do jednego modelu, stanowiąc spójny zasób spełniający wymogi interoperacyjności. Dane zgodne ze specyfikacjami INSPIRE są z założenia zharmonizowane, ponieważ ich modele zostały opracowane zgodnie z zasadami interoperacyjności. Konsekwencją tego, transformacja różnych zbiorów danych o różnych modelach do modeli INSPIRE sprawia, że w wyniku powstają zbiory o modelu zharmonizowanym, co w uproszczeniu można nazwać harmonizacją danych.

2.2. Różnice pomiędzy formatem i językiem

Nowe technologie oparte na języku znacznikowym XML i jego aplikacjach zmieniły fundamentalnie podejście do zapisu danych przestrzennych i do związanych z tym metodami operowania na takich danych. Aby to efektywnie wykorzystać, potrzebne jest zrozumienie różnic jakie są między zapisem za pomocą języka znacznikowego a tradycyjnymi formatami, do niedawna stosowanymi powszechnie.

Tradycyjną formą zapisu danych przestrzennych jest ściśle zdefiniowany sztywny format, który określa semantykę, strukturę logiczną, a także często treść zapisywanych danych. Przykładem takiego zapisu jest format *ESRI Shape File*, którego specyfikacja jest opublikowana i dzięki temu, a także dzięki jego prostocie, stał się on najpopularniejszym sposobem zapisu danych przestrzennych. Specyfikacja tego formatu określa ogólną organizację zbioru danych w postaci kilku plików o wspólnej nazwie i różnych rozszerzeniach. Podstawę formatu stanowią dwa pliki, jeden z zapisem geometrii wyróżnień (obiektów) i drugi w postaci tabeli dBase o stałej, określonej dla danego typu wyróżnień, sekwencji kolumn, zawierający ich atrybuty i o stałej liczbie wierszy równej liczbie wyróżnień w pliku z zapisem ich geometrii. Ze specyfikacji tego formatu wynikają poniższe ograniczenia:

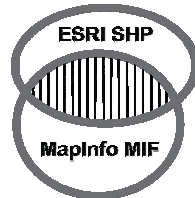
- jeden zestaw danych może opisywać tylko jeden typ wyróżnienia o jednakowej wymiarowości geometrycznej, punkty, krzywe lub powierzchnie, np. odcinki rzek z tabelą atrybutów odpowiednią tylko dla nich i o geometrii w postaci linii;
- wyróżnienia nie mają nazwy typu, i w konsekwencji jako nazwę typu przyjmuje się nazwę zestawu danych, np. zestaw danych o nazwie rzeki zawiera wyróżnienia będące odcinkami rzek i tylko takie wyróżnienia mogą być zapisane w tym zestawie;
- realizacja asocjacji pomiędzy wyróżnieniami jest trudna do implementacji i najczęściej nie obsługiwana przez oprogramowanie przetwarzające dane w tym formacie, ponieważ wykracza to poza ramy określone w jego specyfikacji;
- w całym zestawie danych może być stosowany tylko jeden układ odniesienia;
- zapis nazw atrybutów i ich wartości w tabeli jest ograniczony wymaganiami specyfikacji dBase;
- kodowanie danych tekstowych w tabeli może być oparte na różnych standardach, jednak brak jest informacji jaki sposób kodowania został zastosowany w określonym zestawie danych; w takich przypadkach najczęściej przyjmuje się kodowanie systemowe (przyjęte w danym systemie operacyjnym), np. w systemach Microsoft zależy to od wersji językowej tego systemu – w przypadku wersji polskojęzycznej jest to CP1250.

Podobne ograniczenia występują także w innych sztywnych, prostych formatach. Z tego względu transformacja danych przestrzennych pomiędzy takimi formatami jest najczęściej nieodwracalna, ponieważ może dotyczyć tylko takich form i typów wyróżnień, które dają się zapisać w obu formatach – źródłowym i wynikowym. Można to nazwać w dużym uproszczeniu częścią wspólną obu formatów (rys. 2.3 A).

Szczególnie złożonym problemem jest wyprowadzanie danych z bazy do prostego formatu o ograniczonych pod względem struktury możliwościach. Baza danych ma najczęściej złożony model, składający się z wielu tabel powiązanych ze sobą asocjacjami i nie wszystkie tabele muszą mieć dowiązane elementy geometryczne. W przypadku przedstawionym na rysunku 2.4 tylko pojedyncza tabela wraz z przypisaną geometrią może być wyprowadzona do jednego zbioru w formacie SHP (*ESRI Shape File*). Dla wyprowadzenia pozostałych

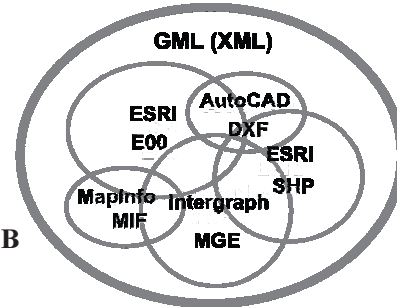
Rys. 2.3. Schematyczne porównanie możliwości transformacji pomiędzy prostymi formatami i językiem GML opartym na XML. Transformacja z jednego prostego formatu na inny zawsze powoduje straty, a zastosowanie języka GML eliminuje ten problem

Transformacja może objąć jedynie część wspólną obu modeli danych



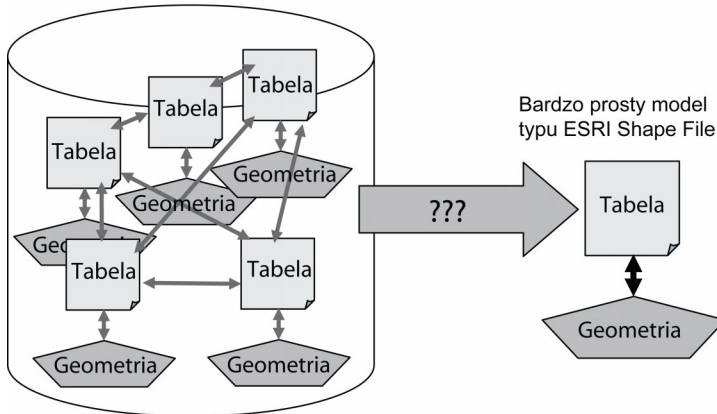
A

GML jest językiem i jego model jest tak obszerny, że obejmuje prawie wszystkie formaty geoinformacji



B

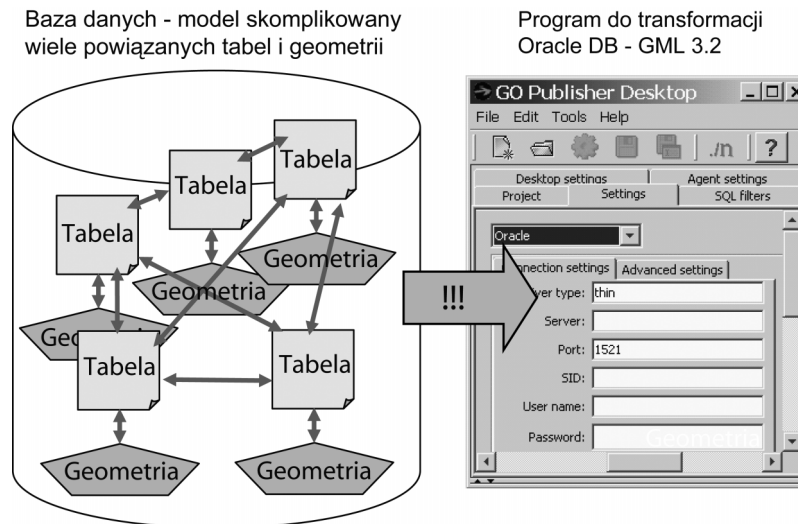
Baza danych - model skomplikowany wiele powiązanych tabel i geometrii



Rys. 2.4. Próbkę danych w prostym formacie (np. *ESRI Shape File*) wyprowadzonych z relacyjnej bazy danych. Czy w takim zbiorze można poprawnie i kompletnie zapisać dane z bazy?

danych z bazy można użyć kolejnych zbiorów tego formatu, jednak zachowanie asocjacji pomiędzy tabelami jest praktycznie bardzo trudne z powodu braku oprogramowania, które mogłoby takie powiązania później odtworzyć.

W przeciwieństwie do prostych sztywnych formatów, języki znacznikowe oparte na XML pozwalają zapisywać różnorodne dane o skomplikowanych zagnieżdżonych strukturach z różnymi wzajemnymi powiązaniem, nie tylko w obrębie jednego zbioru, ale także pomiędzy różnymi zbiorami. W jednym pliku może wystąpić jednocześnie wiele typów elementów o różnej semantyce i strukturze, a elementy te mogą, jeżeli schemat na to pozwala, występować na różnych poziomach hierarchicznej struktury tego pliku. Schematyczne porównanie obszaru możliwości zapisu złożonych struktur danych pomiędzy językiem GML i tradycyjnymi formatami przedstawia rysunek 2.3 B. Te właściwości zapisu znacznikowego XML, zastosowane do danych przestrzennych za pomocą języka GML i jego aplikacji, pozwalają na pełną transformację danych, także w przypadkach, gdy dane źródłowe mają skomplikowaną budowę. Taki przypadek jest przedstawiony na rysunku 2.5, gdzie dane z bazy są wyprowadzane do zapisu aplikacji GML za pomocą oprogramowania GO Publisher firmy Snowflake Software. Bardziej szczegółowy opis tego oprogramowania znajduje się w rozdz.



Rys. 2.5. Wyprowadzanie danych z bazy w formie pliku zapisanego w języku GML wersji 3.2 z zastosowaniem oprogramowania GO Publisher firmy Snowflake Software

2.5. Tu należy zwrócić uwagę na możliwość bezpośredniej komunikacji tego programu z bazą danych zarządzaną przez system Oracle oraz również bezpośrednie wyprowadzanie danych z bazy w zapisie języka GML wersja 3.2, co jest zgodne z wymaganiami specyfikacji INSPIRE. Podobne możliwości zapisu danych przestrzennych w języku GML mają również inne zaawansowane programy przeznaczone do zarządzania danymi i ich transformacji. W pierwszej kolejności trzeba tu wymienić komercyjny system FME (*Feature Manipulation Engine*) i część modułów otwartego systemu Deegree 3, który również jest opisany w rozdz. 2.5.

Przewaga języka GML nad tradycyjnymi formatami zapisu danych przestrzennych polega na tym, że jest to rzeczywiście język, który ma zdefiniowane zasady posługiwania się nim:

- ścisłą syntaktykę określoną w specyfikacji XML,
- rozbudowaną składnię, również określoną w specyfikacji XML,
- zdefiniowane podstawowe typy elementów prostych,
- możliwość definiowania nowych typów elementów prostych na bazie podstawowych,
- określone reguły, przy jednoczesnej swobodzie definiowania elementów złożonych, a także możliwość modyfikowania elementów już wcześniej zdefiniowanych,
- swobodne tworzenie hierarchicznych struktur elementów złożonych,
- zapisywanie wszystkich definicji dotyczących nowej aplikacji w schematach tworzących unikatowy słownik wyróżniony prefiksem przestrzeni nazw,
- zapisywanie tych schematów również w języku XML,
- schematy dotyczące poszczególnych aplikacji mogą tworzyć wspólnie hierarchiczne struktury, które z kolei mogą być użyte do zapisu danych w jednym pliku (dokumencie) aplikacji GML.

Powyższe zasady ustanawiają określony rygor formalny dając jednocześnie wielką swobodę w posługiwaniu się językiem GML. Podstawowe reguły formalne GML (OGC, 2007) wymieniono poniżej.

1. Zgodnie z zasadami XML plik danych przestrzennych zapisanych w języku GML ma strukturę hierarchiczną. Na najwyższym poziomie występuje jeden element typu *FeatureCollection* (rys. 2.19, rozdz. 2.7) – zbiór wyróżnień, którymi są najczęściej obiekty przestrzenne, ale mogą to także być inne zjawiska świata rzeczywistego, jak na przykład obserwacje, zdarzenia, procesy lub obiekty nieprzestrzenne. W wersji 3.2 jako element najwyższego poziomu może także wystąpić inny element złożony, pod warunkiem, że będzie on wyrowadzony z typu *FeatureCollection*. W drugim poziomie hierarchii od góry występują tylko elementy typu *FeatureMember* jako składniki zbioru *FeatureCollection*.
2. Element drugiego poziomu typu *FeatureMember* może mieć opisane odniesienie przestrzenne za pomocą dowolnego zdefiniowanego układu odniesienia. W jednym dokumencie (pliku) można stosować różne układy odniesienia dla różnych elementów tego typu, co w praktyce jest jednak rzadko stosowane.
3. Każdy element drugiego poziomu typu *FeatureMember* ma niepowtarzalny identyfikator, albo w formie parametru (rys. 2.19, rozdz. 2.7), albo w formie elementu prostego (rys. 2.8, rozdz. 2.3), który może być wykorzystany do zapisu powiązań asocjacyjnych pomiędzy wyróżnieniami.

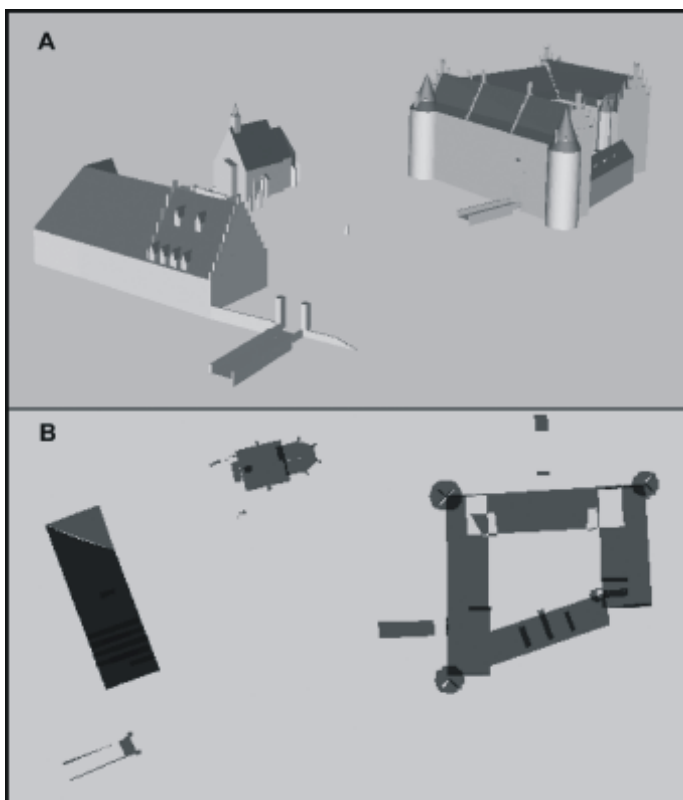
Przykładami aplikacji języka GML stanowiącymi jednocześnie jego rozszerzenie lub przykładami niezależnych rozszerzeń opracowanych w OGC i ISO/TC211, które można wykorzystywać w różnych zastosowaniach, są m.in.:

- SensorML – *Sensor Markup Language* – język przeznaczony do definiowania właściwości sensorów pozyskujących dane przestrzenne;
- TransML – *Transducer Markup Language* – język dla strumieniowego przesyłania danych przestrzennych z sensorów lub przechowywania tych danych;
- GeoXACML – *Geospatial eXtensible Access Control Markup Language* – rozszerzenie języka XACML do zastosowań w geoinformacji; XACML jest językiem przeznaczonym do zapisu praw dostępu do informacji;
- KML – *Keyhole Markup Language* – opracowany dla Google Maps język adnotacji dla miejsc o określonym położeniu; obecnie jest w OGC integrowany z językiem GML;
- CityGML – *City Geography Markup Language* – aplikacja GML dla opisu przestrzeni miejskiej z pełnym wykorzystaniem odniesień trójwymiarowych (rys. 2.6);
- Metadata – *XML schema implementation* – język zdefiniowany w normie ISO 19139 dla zapisu metadanych opisujących dane przestrzenne.

Opisanie pełnych możliwości języka GML wymagałoby odrębnej obszernej publikacji. Tu można zwrócić jedynie uwagę na jego najważniejsze cechy, w porównaniu z formatami tradycyjnymi. Ze względu na uniwersalność i złożoność tego języka, opracowywanie jego aplikacji jest zadaniem trudnym i wymagającym dużej wiedzy i doświadczenia. W tym przypadku mamy do dyspozycji szeroki wachlarz różnych możliwości (Michalak, 2003), od ścisłego trzymania się zaleceń dokumentów specyfikujących ten język, np. normy ISO 19136 i dokumentów określających aplikacje, jakimi są także specyfikacje danych INSPIRE (rys. 2.19, rozdz. 2.7), do opracowywania dowolnych aplikacji XML, w których elementy z przestrzeni nazw GML występują jedynie sporadycznie (rys. 2.22, rozdz. 2.7). W każdym przypadku, a szczególnie gdy odstępstwa od zasad są znaczne, opracowanie aplikacji może sprawić problemy, które w dużym skrócie przedstawione są w następnym rozdziale.

2.3. Problemy implementacji języka GML

Cechą generalną języka GML, podobnie jak to ma miejsce w przypadku obiektowych języków programowania takich jak C++ i Java, jest duża swoboda w posługiwaniu się jego regułami. Opracowanie schematu aplikacyjnego, spełniającego wszystkie wymagania syntaktyczne i składniowe specyfikacji GML i XML, może przynieść różne rezultaty. Może powstać pożyteczna, funkcjonalna i praktyczna aplikacja dla danej dziedziny zastosowań, ale może to być także jej przeciwieństwo. Samo przestrzeganie reguł formalnych języków znacznikowych nie gwarantuje sukcesu. Przyczyną nieudanej aplikacji może być źle opracowany model pojęciowy danych przestrzennych z tej dziedziny. Takie przypadki występują, gdy nowo opracowany model jest wiernym obrazem struktur i form stosowanych do zapisu danych w przeszłości. Inną przyczyną może być nieznanostwo specyfiki języków znacznikowych opartych na specyfikacji XML. Konsekwencją tego może być definiowanie niepotrzebnych typów elementów lub zbędne w wielu przypadkach, przekształcanie już zdefiniowanych typów elementów, w celu dostosowania ich do z góry przyjętej wizji autora aplikacji. Można tu użyć prostej metafory, że języki znacznikowe są podobne do języków naturalnych i w obu przypadkach przestrzegając zasad formalnych można się nimi posługiwać dobrze lub źle.



Rys. 2.6. Zobrazowanie zapisu aplikacji CityGML: A – za pomocą przeglądarki Arystoteles3D z pełnym wykorzystaniem możliwości aplikacji GML w zakresie trzech wymiarów, B – za pomocą przeglądarki Snowflake GO GML Viewer, której interpretacja jest ograniczona do dwóch wymiarów w planie (Michalak, 2008)

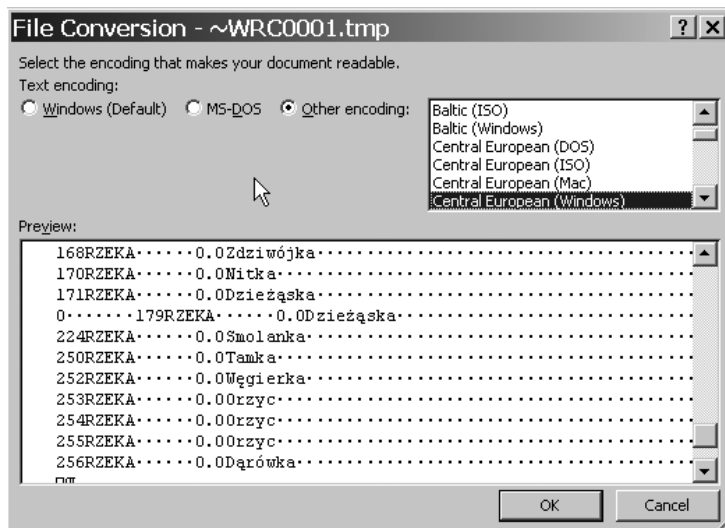
Innym ważnym problemem aplikacji GML jest ich implementacja w postaci oprogramowania, które może tworzyć dokumenty (pliki) zgodne ze schematem aplikacyjnym lub te dokumenty interpretować w celu dalszego przetwarzania lub zamiany na obraz graficzny dla zobrazowania zapisanych w tych dokumentach danych przestrzennych. Takie programy analizują zawartość dokumentu i wyszukują elementy, które są im „znane”, a następnie podejmują określoną akcję, gdy napotkają znany im element. W przypadku napotkania nieznanego elementu lub nieznanego składnika znanego elementu program analizujący dokument może podjąć różne akcje wyjątkowe. Jedną z takich akcji jest „ciche” pominięcie nieznanego elementu lub składnika. Oznacza to, że program-interpretator nie informuje użytkownika o tym, że znalazł nieznaną element i taka akcja jest traktowana jako standardowa akcja normalna. Większość przeglądarek języka GML tak reaguje na nieznaną elementy lub składniki wprowadzone do schematu aplikacyjnego. Innym przypadkiem, w którym pomija się elementy lub ich składniki, jest niepełna implementacja podstawowej części specyfikacji języka GML. Przykładem tego może być fakt, że w języku GML jest zdefiniowanych wiele form geometrycznych o różnorodnej wymiarowości – od punktu (0D) w przestrzeni 1D do skomplikowanej bryły (3D) w przestrzeni 3D. Jednak zdecydowana większość przeglądarek GML ogranicza interpretację zapisów do płaskich współrzędnych x i y (przestrzeń 2D), celowo pomijając trzeci wymiar z . W takim przypadku pomijane składniki są znane przeglądarce, jednak ze względu na jej ograniczenia są przez nią „świadomie” pomijane. Przykład tego przedstawiony jest na rysunku 2.6.

Z powyższych uwag wynika wniosek, że ograniczenie się do pierwszych dwóch faz opracowania aplikacji języka GML, do opracowania modelu pojęciowego w UML i przekształcenie go w schemat XSD, bez ostatniej fazy aplikacyjnej tworzącej odpowiednie oprogramowanie, jest działaniem nieracjonalnym (Carlson, 2001). Praktyczne zastosowanie aplikacji języka GML wymaga użycia niezbędnego oprogramowania, które będzie potrafiło poprawnie interpretować wszystkie elementy i ich składniki zdefiniowane w tej aplikacji.

Dla zilustrowania problematyki implementacji języka GML, szczególnie w przypadku transformacji danych z formatów tradycyjnych do nowo opracowanego schematu, można przytoczyć trzy wybrane przykłady.

Przykład 1. Kodowanie tekstów. W pojedynczym pliku zapisu XML kodowanie tekstów jest określone w atrybucie parametrycznym nagłówka pliku, np. dla 8-bitowego Unicode w postaci: `<?xml version="1.0" encoding="UTF-8"?>` lub z inną wartością tego atrybutu dla innego sposobu kodowania. W konsekwencji wszystkie przeniesione do tego pliku dane z innych zapisów muszą mieć teksty kodowane zgodnie z tą deklaracją. W przeciwnym razie błędy kodowania, które powstaną z przyczyny niezgodności kodów, nie będą mogły być poprawione po przeniesieniu do zapisu XML. W przypadku prostych tradycyjnych formatów, które są źródłem danych dla nowego zapisu w języku GML, ustalenie sposobu kodowania tekstów nie jest zadaniem prostym. Praktyczne rozwiązanie tego problemu przedstawia rysunek 2.7.

Przykład 2. Asocjacje. Przedstawiony jest problem realizacji asocjacji pomiędzy obiektami klas języka UML w dokumentach XML (w tym przypadku w dokumentach aplikacji GML) za pomocą powiązań pomiędzy elementami. Mechanizm wiązania wykorzystujący odsyłacze `xlink:href` realizuje takie asocjacje i to nie tylko wewnątrz jednego pliku, ale także pomiędzy różnymi plikami, również znajdującymi się w różnych repozytoriach. Jednak mechanizm ten realizuje jedynie powiązanie jednokierunkowe (z nawigacją w jedną stronę). Zrealizowanie nawigacji dwukierunkowej wymaga użycia dwóch odsyłaczy komplementarnych i skierowanych w przeciwne strony, jak to jest przedstawione na rysunku 2.8.



Rys. 2.7. W przypadku braku informacji o kodowaniu tekstów w zbiorze danych przestrzennych ustalenie standardu nie jest zadaniem prostym. W wielu przypadkach potrzebna jest metoda próbkowania – można do tego użyć programu MS Word otwierając za jego pomocą dowolny plik zawierający teksty



Rys. 2.8. Przykład wzajemnego wiązania elementów zapisu danych w języku GML z wykorzystaniem wbudowanego w ten język mechanizmu *xlink:href* – jednego z wielu rozszerzeń XML.

Przykład ten jest fragmentem zapisu opartego na schemacie aplikacyjnym tematu *obszary chronione* dyrektywy INSPIRE i dotyczy powiązań pomiędzy miejscami chronionymi i instytucjami dbającymi o ich ochronę

Przykład 3. Interpretacja różnych sposobów zapisu współrzędnych. W prostych formatach stosowanych wcześniej zapis współrzędnych był ściśle określony. Z reguły były to dwie współrzędne układu płaskiego, określane ogólnie jako *easting* i *northing*, bez wnikania czy to jest to oś *x* czy *y*, co w różnych układach odniesienia jest przyjmowane różnie. Uniwersalność języka GML sprawiła, że kolejność i sposób zapisu współrzędnych stworzyły pewien problem, spotęgowany dodatkowo różnicami w tym zakresie pomiędzy wersją 2 i 3. W rezultacie w wersji 3 współrzędne można zapisać za pomocą elementu *gml:pos* lub *gml:posList* i oba zapisy mogą być różnie interpretowane przez oprogramowanie czytające zapis GML, także przez przeglądarki (rys. 2.9). W wielu programach brak jest domyślnej kolejności współrzędnych przy interpretacji zapisu, a nieliczne pozwalają na wybór z określonej listy, np. pakiet FME daje możliwość wyboru pomiędzy czterema wariantami kolejności: „1, 2”, „2, 1”, „1, 2, 3” i „2, 1, 3”.

Generalny wniosek, jaki wynika z przedstawionych powyżej faktów i analizy ich wzajemnych powiązań, można wyrazić następująco – skuteczne, poprawne i praktyczne zastosowanie języka GML do zapisu danych przestrzennych w określonej dziedzinie jest w pierwszej kolejności uzależnione od właściwości oprogramowania, które będzie zapisy tej aplikacji generowało lub wczytywało i analizowało. Aplikacja GML – mając poprawnie opracowany model pojęciowy w UML oraz poprawne schematy XSD zgodne z tym modelem – bez odpowiedniego oprogramowania interpretującego wszystkie elementy tych schematów jest praktycznie bezużyteczna.

2.4. Specyfikacje danych INSPIRE

Wymagania INSPIRE, dotyczące struktur i treści danych przestrzennych, są określone w specyfikacjach danych w formie dokumentów tekstowych. Jednak ich precyzyjny zapis jest zawarty w modelach danych zapisanych w diagramach klas języka UML i schematach aplikacyjnych XSD języka GML. Obecnie (stan na wrzesień 2010 r.) lista tych dokumentów składa się z 24 pozycji, które można pogrupować następująco:

- raporty o charakterze instrukcji dotyczące metodyki opracowywania specyfikacji danych (INSPIRE, 2007; 2008a; 2008c; 2008f; 2008g);
- robocze wersje przepisów implementacyjnych z zakresu transformacji (INSPIRE, 2009a; 2009b; 2010a);
- modele w języku UML i schematy XSD dla tematów z załącznika I dyrektywy INSPIRE (INSPIRE, 2010b; 2010d);
- specyfikacje danych dla tematów z załącznika I dyrektywy INSPIRE – jednostki administracyjne, działki ewidencyjne, nazwy geograficzne, hydrografia, obszary chronione, sieci transportowe, adresy (INSPIRE, 2010e; 2010f; 2010g; 2010h; 2010i; 2010j; 2010k);
- specyfikacje układów odniesienia i siatek – instrukcje techniczne (INSPIRE, 2010m; 2010n);
- dokumenty ogólne w formie raportów i repozytoriów plików wspomagające proces opracowywania specyfikacji i realizacji ich ustaleń (INSPIRE, 2008b; 2008d; 2008e; 2010c; 2010l);

Analiza tych dokumentów pozwala na sformułowanie kilku ogólnych uwag:

1. Do zapisu danych ma być stosowany wyłącznie język GML w wersji 3.2.1, jednak ze względu na to, że obecnie bardzo mało systemów obsługuje tę wersję poprawnie, dopuszcza się przejściowe stosowanie wersji 3.1.1.
2. Schematy XSD dla tematów INSPIRE są wyjątkowo skomplikowane w porównaniu z innymi podobnymi aplikacjami tego rodzaju, np. brytyjskim Master Map lub niemieckim AAA-NAS.
3. Skomplikowanie to wynika m.in. z bardzo wielu powiązań asocjacyjnych pomiędzy elementami, nie tylko w obrębie jednego dokumentu lub tematu, ale także pomiędzy różnymi tematami.
4. Powiązania te są realizowane za pomocą mechanizmu *xlink:href*. Dla zrealizowania asocjacji dwukierunkowych potrzebne są dwa takie odsyłacze komplementarne skierowane w przeciwnych kierunkach (rys. 2.8, rozdz. 2.3).
5. W schematach XSD, obok standardowych identyfikatorów GML, definicje typów elementów określa wiele innych różnych identyfikatorów do różnych celów, co komplikuje sposób odwoływania się do poszczególnych elementów zapisu.

6. Elementy złożone składają się z kolejnych elementów złożonych i często to zagnieżdżanie jest wielostopniowe. Ta wielopoziomowość struktur elementów stwarza duże trudności w obsłudze zapisów danych przez oprogramowanie do tego celu przeznaczone.
7. Wszystkie dotychczas opracowane schematy dotyczą wyróżnień (obiektów przestrzennych) zapisanych wektorowo. Można oczekiwać, że dopiero w dalszych tematach wystąpią inne formy danych, np. pokrycia w różnych sposobach zapisu odpowiednich do treści (Herring, Kottman, 1997; Mark, Skupin, Smith, 2001; Smith, 2001; Michalak, 2005).

Przykład 4. Zapis struktury wielopoziomowej. Ilustracja problemów zapisu struktury wielopoziomowej, która ze względu na ograniczenia oprogramowania musiała być spłaszczona, przez użycie wielocłonowych nazw elementów niższych poziomów.

```
<gml:featureMember>
  <fme:NamedPlace gml:id="id8e490696-5f3f-44c3-932c-fbad36da89fb">
    <fme:name_GeographicalName_language>Italian
  </fme:name_GeographicalName_language>
    <fme:name_GeographicalName_nativeValue>MODICA
  </fme:name_GeographicalName_nativeValue>
    <fme:type>Populated place</fme:type>
    <fme:levelOfDetail>3</fme:levelOfDetail>
    <fme:typeLocal>Regione</fme:typeLocal>
    <gml:pointProperty>
      <gml:Point srsName="EPSG:4326" srsDimension="2">
        <gml:pos>36.8639984130859 14.7620000839233</gml:pos>
      </gml:Point>
    </gml:pointProperty>
  </fme:NamedPlace>
</gml:featureMember>
```

Zapis powyższego fragmentu, z uwzględnieniem wielopoziomowości zgodnie z wymaganiami specyfikacji, jest następujący:

```
<gml:featureMember>
  <GN:NamedPlace gml:id="id8e490696-5f3f-44c3-932c-fbad36da89fb">
    <GN:name>
      <GN:GeographicalName>
        <GN:language>Italian</GN:language>
        <GN:nativeValue > MODICA </GN:nativeValue >
      </GN:GeographicalName>
    </GN:name>
    <GN:type>Populated place</GN:type>
    <GN:levelOfDetail>3</GN:levelOfDetail>
    <GN:typeLocal>Regione</GN:typeLocal>
    <gml:pointProperty>
      <gml:Point srsName="EPSG:4326" srsDimension="2">
        <gml:pos>36.8639984130859 14.7620000839233</gml:pos>
      </gml:Point>
    </gml:pointProperty>
  </GN:NamedPlace>
</gml:featureMember>
```

Różnica pomiędzy tymi dwoma zapisami polega na przeniesieniu elementów `<GN:language>` i `<GN:nativeValue>` o dwa poziomy wyżej i zmianie ich nazw na `<fme:name_GeographicalName_language>` i `<fme:name_GeographicalName_nativeValue>`. W rezultacie tego elementy `<GN:name>` i `<GN:GeographicalName>` nie są potrzebne. Taki spłaszczony zapis nie jest zgodny ze specyfikacjami, ale może być w dalszej kolejności przekształcony do poprawnej formy przez zastosowanie odpowiedniego oprogramowania.

Prace studialne i eksperymentalne, dotyczące poprawnych zapisów danych przestrzennych zgodnie ze specyfikacjami danych INSPIRE, są żmudne. Wymagają modyfikacji oprogramowania, które jest do tych prac stosowane. W następnym rozdziale (2.5) przedstawione będą trzy programy, które mogą służyć do takich zadań – w szczególności do prac z zakresu transformacji z istniejących polskich zbiorów danych o różnych modelach, różnych formach zapisu i różnych zestawach danych atrybutowych. W tym przypadku istotną przeszkodą jest brak przykładów danych zapisanych zgodnie z wymaganiami specyfikacji INSPIRE, nawet w postaci niewielkich próbek, które są niezbędne do prac nad przystosowaniem oprogramowania to takich potrzeb.

2.5. Oprogramowanie dedykowane transformacji

Wiele różnych systemów programowych, poczynając od prostych programów typu desktop, kończąc na wyspecjalizowanym oprogramowaniu dla geoserwerów, ma możliwości zmiany formy zapisu danych przestrzennych. Jednak w przypadku wymagań stawianych danym INSPIRE, wybór odpowiedniego oprogramowania nie jest zadaniem prostym. W dużym skrócie wymagania te można określić następująco:

1. Docelową formą zapisu jest GML wersji 3.2.1, ale tymczasowo może to być wersja 3.1.1.
2. Struktura danych docelowych jest określona w schematach XSD zdefiniowanych dla każdego tematu INSPIRE.
3. Oprogramowanie powinno umożliwić budowanie struktur GML z wielopoziomym zagnieżdżeniem elementów i tworzeniem powiązań asocjacyjnych w plikach wynikowych, także w przypadkach gdy dane źródłowe tego nie mają.
4. Na obecnym etapie prac nad dokumentami INSPIRE nie ma jeszcze określonej ogólnej organizacji danych na poziomie plików. Można przypuszczać, że będzie w tym zakresie duża swoboda i prawdopodobnie poszczególne pliki będą zawierały różne typy wyróżnień, zgrupowane razem w obrębie tematu i w ewentualnym podziale na obszary, tak jak jest to przyjmowane najczęściej w innych zastosowaniach języka GML.
5. W zdecydowanej większości przypadków modele danych INSPIRE znacznie się różnią od obecnych danych polskich odpowiadających swoim zakresem tematycznym danym INSPIRE.
6. Dla utworzenia jednego zbioru danych INSPIRE trzeba będzie wykorzystać więcej niż jeden zbiór danych polskich, a w pewnych przypadkach może być ich kilka, co stwarza trudną sytuację jednoczesnej transformacji kilku zbiorów źródłowych.
7. Konieczność przeprowadzenia znaczącej zmiany organizacji danych wymaga zastosowania zaawansowanych technologii opartych na XSLT. Oprogramowanie zastosowane do transformacji powinno tę technologię wykorzystywać.

8. Potrzebna jest także możliwość zmiany sposobu kodowania tekstów, przeliczania współrzędnych pomiędzy różnymi układami odniesienia, a także w pewnych przypadkach, zamiany kolejności współrzędnych.

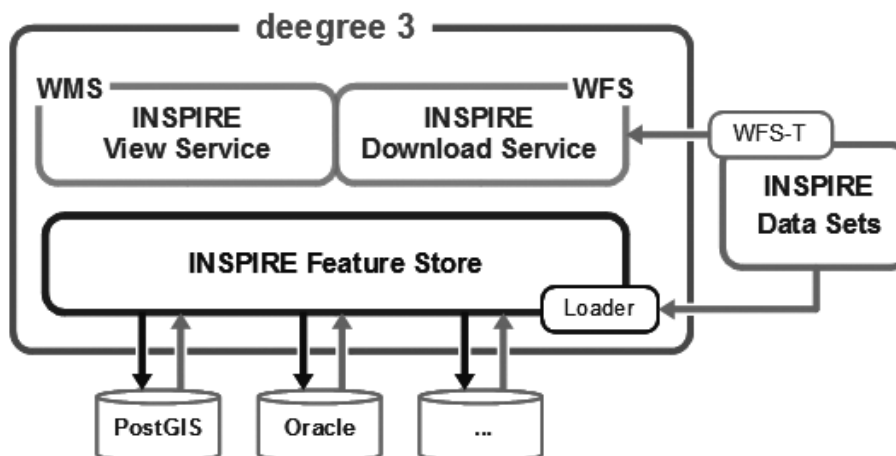
Przegląd oprogramowania pozwolił na wytypowanie trzech systemów, które najpełniej spełniają powyższe wymagania. Jednak żaden z tych systemów nie spełnia wszystkich wymagań. Są także inne systemy dedykowane zadaniom wynikającym z dyrektywy INSPIRE. W tej grupie można wymienić dwa przykładowe: 1) ArcGIS SDI platforma firmy ESRI (ESRI, 2010), 2) Geomedia SDI pro firmy Intergraph (Intergraph, 2010). Jednak z powodu braku bardziej szczegółowych informacji o ich możliwościach i zaletach w zastosowaniach do transformacji systemy te nie były rozpatrywane.

Trzy wybrane w wyniku przeprowadzonej analizy systemy przedstawiono poniżej.

Feature Manipulation Engine (FME) firmy Safe Software (rys. 2.10) w wersji desktop dla transformacji *off-line* i w wersji serwer dla transformacji *on-line*. Obie wersje są dostępne dla platform: MS Windows, Linux Red Hat, IBM AIX i Sun Sparc Solaris. Oprogramowanie to jest obecnie wiodącym na świecie systemem dedykowanym zadaniom transformacji danych przestrzennych, obsługuje ponad 250 różnych formatów, interfejsów i języków, a także ma wbudowany procesor XSLT. Pewnym ograniczeniem tego oprogramowania jest nieprzystosowanie go do wymagań INSPIRE, jednak są prowadzone prace w tym zakresie (Safe, 2010).

System GO Publisher firmy Snowflake Software (rys. 2.11) dedykowany wyprawdzaniu danych z bazy zarządzanej systemem Oracle do plików GML zgodnych z wymaganiami specyfikacji INSPIRE (Snowflake, 2010). Jest to komponent pakietu, w którego skład wchodzi także oprogramowanie GO Loader dla wprowadzania danych w GML do bazy zarządzanej przez Oracle i Go Viewer – bezpłatna przeglądarka danych GML wersji 2, 3.1 i 3.2, dostosowywana w miarę potrzeb do różnych schematów aplikacyjnych, w tym brytyjski Master Map, niemiecki AAA-NAS i holenderski TOP10NL. GO Publisher jest dostępny w dwóch wersjach – komercyjnej i tzw. *Community Edition* – bezpłatnej. Obecnie trwają prace nad dostosowaniem tego oprogramowania do współpracy z otwartym systemem PostgreSQL i jego rozszerzeniem PostGIS. GO Publisher może pracować na platformach sprzętowo-systemowych MS Windows, Linux i Sun Sparc Solaris.

Deegree 3. W przeciwieństwie do dwóch ww. systemów komercyjnych, system ten należy do kategorii oprogramowania otwartego. Jest to nowa wersja znanego już wcześniej oprogramowania Deegree (Schneider, 2010), szczególnie dedykowana zadaniom INSPIRE (rys. 2.12, 2.13 i 2.14). Jest to oprogramowanie rozwijane w technologii języka Java, co ma swoje zalety, ale także wady. Do zalet można zaliczyć niezależność od platformy sprzętowo-systemowej, obiektowość i wielką liczbę otwartych bibliotek, co przekłada się na łatwość budowy oprogramowania. Mankamentem jest to, że oprogramowanie w języku Java działa w środowisku maszyny wirtualnej, jest kompilowane do przenośnego kodu, który następnie jest w maszynie wirtualnej kompilowany w trakcie realizacji (*runtime compilation*) lub tuż przed (*just-in-time compilation*). W porównaniu do kodu binarnego, każdy uruchomiony nowy proces na serwerze trwa dłużej lub jest wyraźnie opóźniony, przez konieczność wykonania kompilacji w maszynie wirtualnej. Sposobem na rozwiązanie tego problemu może być: 1) zastosowanie interfejsu FastCGI zamiast zwykłego CGI dla utrzymania uruchomionych procesów w stanie oczekiwania, 2) zastosowanie kompilatora języka Java GCJ (*GNU Compiler for Java*) dla utworzenia kodu binarnego, który nie wymaga użycia środowiska wirtualnego i może być także zastosowany w połączeniu z interfejsem FastCGI.



Rys. 2.14. Ogólny schemat architektury systemu Deegree 3 dedykowanego zastosowaniom w infrastrukturze INSPIRE (cytat z publikacji internetowej: Schneider, 2010)

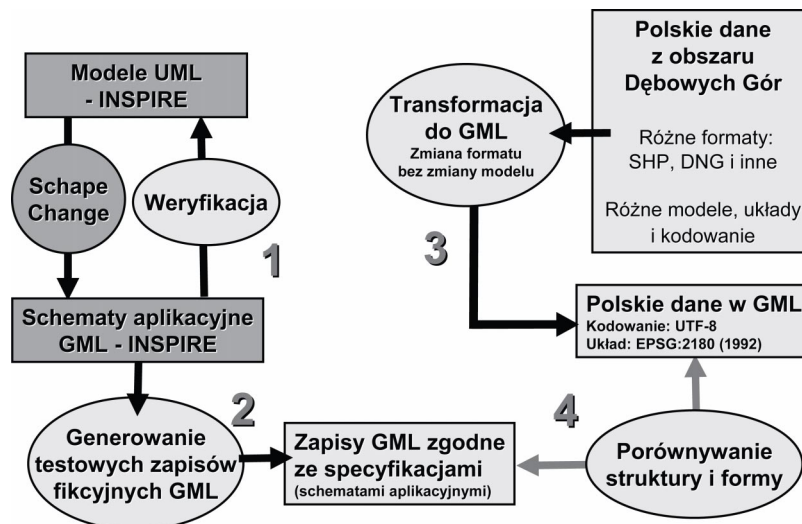
2.6. Testowanie metod transformacji

Praktyczne rozwiązanie problemów transformacji polskich danych do schematów INSPIRE wymagało przeprowadzenia wstępnych prac testowych. Zostały one wykonane w Laboratorium Modelowania i Geomatyki na Wydziale Geologii Uniwersytetu Warszawskiego, w ramach prac powołanego przez Głównego Geodetę Kraju zespołu do testowania specyfikacji danych I grupy tematycznej INSPIRE. Prace były ukierunkowane na problemy technologiczne transformacji danych. Zagadnieniami kompletności polskich danych, w relacji do zawartości danych określonych w specyfikacjach INSPIRE, zajmował się równoległe zespół Instytutu Geodezji i Kartografii. Wyniki tych prac są przedstawione w raporcie pt. „Prace nad identyfikacją zbiorów i usług danych przestrzennych dla I i II grupy tematycznej INSPIRE” (IGiK, 2009).

Ogólny schemat procesu transformacji polskich danych do schematów INSPIRE, a w konsekwencji schemat prac testowych, które podzielone były na 4 fazy, jest przedstawiony na rysunku 2.15.

Faza 1 testowania transformacji. Model danych w języku UML (*Unified Modeling Language*) (Larman, 2001), zapisany w formie XMI (*XML Metadata Interchange*), może być następnie przekonwertowany do schematu aplikacyjnego XSD języka GML. W rezultacie stanowi podstawę do tworzenia plików GML zawierających zbiory danych przestrzennych z wybranego zakresu tematycznego INSPIRE. W zespołach opracowujących specyfikacje danych INSPIRE do przekształcenia modelu UML do schematu XSD stosowany był program Shape Change. W ramach pierwszej fazy prac testowych transformacji, dokonana została operacja odwrotna (*reverse engineering*) – ponowne odtworzenie modelu UML z plików XSD za pomocą programów Mapforce i Umodel firmy Altova oraz porównanie wyniku z modelem początkowym. Test ten wykazał niepełną odwracalność przekształcenia za pomocą programu Shape Change (rys. 2.15, 1).

Faza 2 testowania transformacji. Celem tej fazy (rys. 2.15, 2) było wygenerowanie fikcyjnych zbiorów danych przestrzennych zawierających dane całkowicie niepoprawne pod



Rys. 2.15. Diagram przedstawiający schematyczny algorytm prac testowych z zakresu transformacji polskich danych przestrzennych do schematów INSPIRE (objaśnienia w tekście)

względem treści, ale w pełni zgodne ze schematami XSD INSPIRE dla poszczególnych tematów I grupy. Pozwoliło to na analizę struktury zapisu plików zgodnych z tymi schematami. W przyszłości może być podstawą do utworzenia próbek danych reprezentujących poszczególne typy wyróżnień, jakie w tych zbiorach powinny wystąpić. Wynik fazy drugiej pozwolił określić, jaką strukturę powinny mieć zbiory po transformacji.

Faza 3 testowania transformacji. W tej fazie dokonano transformacji wybranych próbek polskich danych z formatów, w jakich są obecnie dostępne do zapisów w języku GML, jednak z zachowaniem ich obecnych modeli niezgodnych z wymaganiami INSPIRE. Ważnym elementem tej fazy prac było ujednoczenie sposobów kodowania danych, w szczególności tekstowych do kodu UTF-8, a także przeliczenie wszystkich współrzędnych do jednego układu odniesienia EPSG:2180 (Poland CS 92). Proces transformacji danych, z wykorzystaniem w tym przypadku oprogramowania FME, wymagał określenia struktury i treści zarówno danych źródłowych (*Source Types*), jak i danych wynikowych (*Destination Types*). Dysponując tymi dwoma modelami można określić mapowanie (powiązania) poszczególnych elementów danych źródłowych z odpowiadającymi im elementami danych wynikowych. Rysunek 2.16 przedstawia mapowanie z kilku zbiorów polskich dotyczących hydrografii w jeden zbiór GML. W trakcie mapowania można użyć tzw. transformerów dla przeliczania mapowanych wielkości (rys. 2.17), a także dokonujących bardzo wielu innych operacji na elementach i ich atrybutach.

W przypadku, gdy mapowane modele danych znacznie się między sobą różnią, jedynie niewielka część elementów obu modeli może być połączona. Taka sytuacja występuje bardzo często w przypadku mapowania polskich zbiorów danych do modeli INSPIRE. Rysunek 2.16 przedstawia porównanie liczby elementów i ich atrybutów pomiędzy modelem testowanej próbki danych polskich a modelem INSPIRE dla tematu jednostki administracyjne. Dane polskie, wyprowadzone z bazy danych PRG w formacie SHP, obok identyfikatora i geometrii

GML, zawierają jedynie dwa atrybuty OBSZAR i NAZWA, podczas gdy model INSPIRE zawiera ich dziesiątki. Można przypuszczać, że jest to bardzo zubożona próbka polskich danych w wyniku przenoszenia pomiędzy bazami i plikami o różnych formatach.

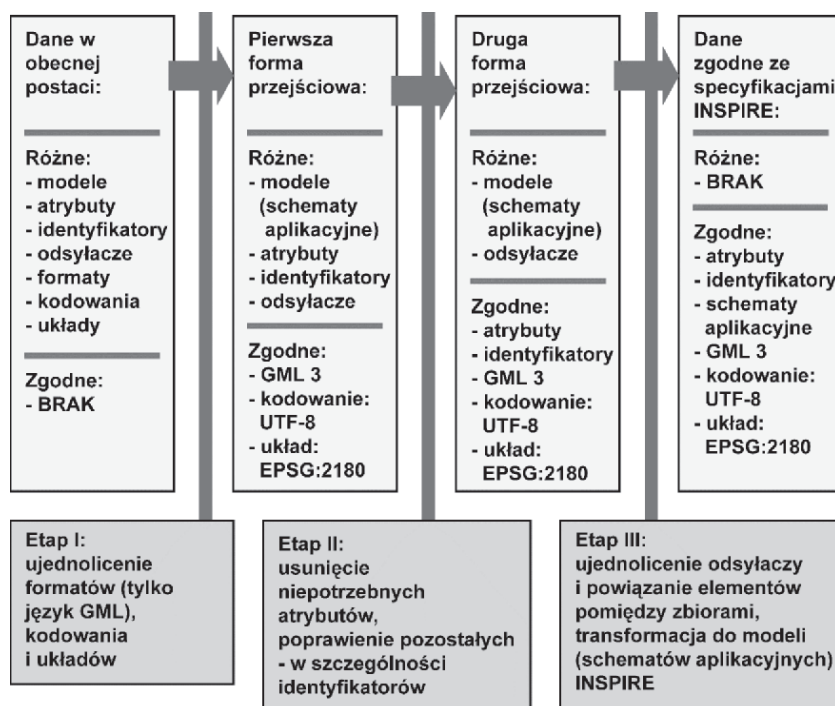
Faza 4 testowania transformacji. Celem prac testowych w tej fazie było porównanie wyników fazy drugiej (fikcyjne zapisy GML zgodne ze specyfikacjami INSPIRE) z wynikami fazy trzeciej (dane polskie w GML z własnymi modelami, ujednolicone jedynie formalnie). Porównanie to wykazało możliwość „ręcznego” transformowania małych próbek danych. Jednak w przypadku transformowania licznych dużych zbiorów wykazało konieczność zastosowania technologii opartej na XSLT oraz dało wiele wskazówek, w jaki sposób tę technologię należy zastosować.

W wyniku przeprowadzonych prac testowych został opracowany ogólny algorytm transformacji (rys. 2.18), który dzieli się na trzy etapy:

Etap I. Transformacja polskich danych do zapisów języka GML z zachowaniem początkowych modeli danych, ale ze zmianą kodowania na UTF-8 oraz przeliczenie współrzędnych na wybrany wspólny układ, np. EPSG:2180 (Poland CS 92).

Etap II. Usunięcie niepotrzebnych atrybutów, poprawienie pozostałych, w szczególności identyfikatorów elementów i wyróżnień (*feature*) oraz przygotowanie danych do utworzenia powiązań za pomocą *xlink:href*.

Etap III. Połączenie wielu zbiorów z zakresu jednego tematu w jeden zgodny ze specyfikacjami INSPIRE, w tym utworzenie powiązań pomiędzy elementami w obrębie jednego tematu. Pozostaje jeszcze utworzenie powiązań pomiędzy tematami, ale ta operacja wymaga



Rys. 2.18. Podział procesu transformacji polskich danych przestrzennych do schematów INSPIRE na trzy etapy (objaśnienia w tekście)

istnienia kompletu danych w obrębie tematów powiązanych ze sobą oraz ustalenia, które zbiory są zależne od innych i w jaki sposób. Z tego względu zadanie to powinno być odłożone na przyszłość.

2.7. Zbiory danych testowych

W ramach prac testowych pozyskano dane dla obszaru poligonu testowego Dębowe Góry (*OakHills*) położonego na północy Mazowsza. Obszar ten obejmuje około 273 km², a w przypadku mniej szczegółowych danych 4 arkusze mapy w skali 1:50 000, co odpowiada w przybliżeniu powierzchni 1600 km². W przypadku braku odpowiednich danych testowych z tego poligonu, zostały pozyskane dane z innych położonych w pobliżu miejsc. Zbiory danych udostępniły następujące instytucje:

Generalna Dyrekcja Ochrony Środowiska

- dane cyfrowe dotyczące obszarów chronionych w formacie SHP,
- dane cyfrowe Mapy Ostoi Przyrody w Polsce w skali 1:50 000 w formacie MDB,

Centralny Ośrodek Dokumentacji Geodezyjnej i Kartograficznej

- dane cyfrowe Bazy Danych Topograficznych w skali 1:10 000 w formacie TBDGML,
- dane cyfrowe VMap Level 2 w skali 1:50 000 w formatach SHP, MDB i TBDGML,
- dane z Państwowego Rejestru Granic i Powierzchni Jednostek Podziałów Terytorialnych Kraju w formacie SHP,
- dane z Państwowego Rejestru Nazw Geograficznych w formacie SHP,
- granice działek ewidencyjnych w formacie GML,

Mazowiecki Ośrodek Dokumentacji Geodezyjnej i Kartograficznej

- dane cyfrowe Mapy Topograficznej w skali 1:10 000 w formacie DNG,

Państwowy Instytut Geologiczny

- dane cyfrowe MHP (Mapy Hydrograficznej Polski) GUGiK w skali 1:50 000 w formacie MIF,
- dane cyfrowe MSP (Mapy Sozologicznej Polski) GUGiK w skali 1:50 000 w formacie MIF,

Krajowy Zarząd Gospodarki Wodnej

- dane cyfrowe MPHP (Mapy Podziału Hydrograficznego Polski) w skali 1:50 000 w formacie SHP,

Powiatowy Ośrodek Dokumentacji Geodezyjnej i Kartograficznej w Przasnyszu

- granice działek ewidencyjnych w formacie DXF,

Nadleśnictwo Przasnysz

- dane cyfrowe LMN (Leśnej mapy numerycznej) w formacie SHP.

Pozyskanie tak dużej liczby różnorodnych zbiorów danych było możliwe dzięki: 1) współpracy z innymi instytucjami w ramach wspólnych prac testowych nad transformacją polskich danych do specyfikacji określonych w dokumentach dyrektywy INSPIRE, 2) życzliwemu podejściu instytucji będących dysponentami tych danych do prac testowych w ramach przygotowań do realizacji zadań wynikających z dyrektywy INSPIRE. Wymienione powyżej zbiory są właściwie „próbkami danych” i są bardzo różnorodne pod względem:

- pochodzenia – samodzielne zbiory, dane wyprowadzone z baz danych do plików, mapy cyfrowe w podziale arkuszowym, często w postaci wielu plików;

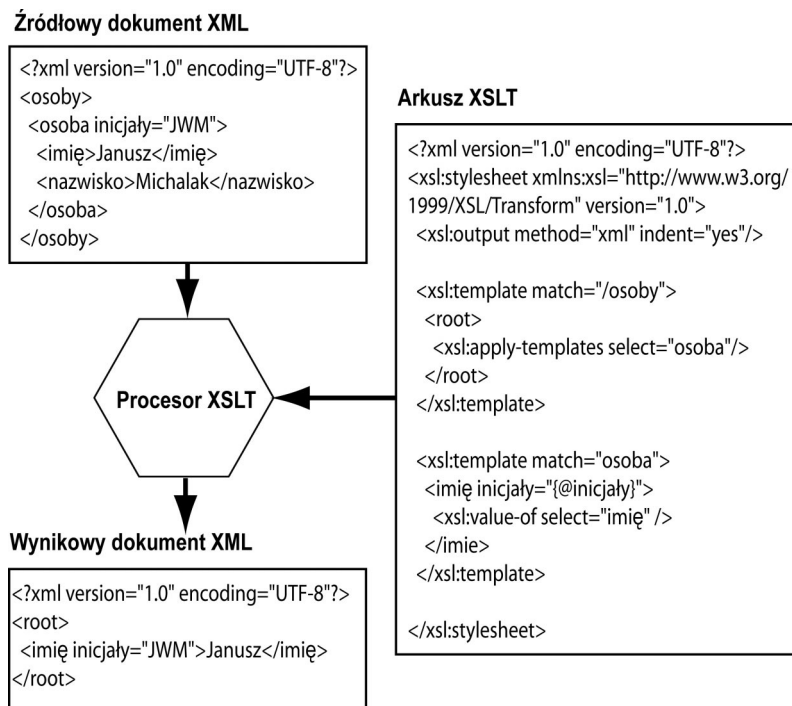
- formy zapisu – różne formaty: SHP, DXF, MIF, pliki baz MDB, zapisy w języku GML i w językach do niego zbliżonych, dokumenty w formatach: TXT, PDF, XLS i inne;
- stosowanych układów odniesienia, np. EPSG:2180 (układ 1992), EPSG:3120,2172-2175 (układ 1965), EPSG:4326 (układ WGS84-LatLong);
- sposobów kodowania tekstów: DOS, CP1250, ISO 8859-2, UTF-8;
- obszaru – część danych była w podziale na arkusze map, a inne obejmowały dowolnie wybrane obszary;
- zapisu danych – w jednym pliku lub w bardzo wielu np. VMap i TBD;
- stopnia szczegółowości i dokładności najczęściej wyrażonego skalą mapy;
- złożoności modelu danych, np. prosty model formatu SHP i złożona struktura bazy w formacie MDB.

Ten bardzo różnorodny materiał poddany został szczegółowym testom, których schemat i algorytm przedstawione były w poprzednim rozdziale (2.6). Podstawowe problemy, na jakie napotkano podczas prac testowych, to brak modeli danych dla testowanych zbiorów i często trudne do ustalenia pochodzenie danych zawartych w tych zbiorach. Problem braku modeli jest w wielu przypadkach zrozumiały, ponieważ dane często były utworzone dość dawno i nie zawsze jest dostępna ich pełna dokumentacja. Dla potrzeb testowania modele były odtwarzane na podstawie zawartości testowanych zbiorów. Jednak dla transformacji pełnych zasobów takie rozwiązanie nie może być przyjęte, zatem opracowanie modeli na podstawie dokumentacji jest koniecznością. W przypadku prac testowych trudności z ustaleniem pochodzenia danych w zbiorach nie mają istotnego znaczenia. Przykładowo dane VMap 2 były testowane we wszystkich trzech dostępnych formach: SHP, MDB i TBDGML, bez ustalania, który z tych zbiorów był pierwotny, a który wtórny. W przypadku transformacji pełnych zasobów to również nie może być pominięte, ponieważ nie ma pewności czy zmiana formy zapisu nie spowodowała zubożenia zawartych w zbiorach informacji.

Rysunki 2.19, 2.20, 2.21 i 2.22 przedstawiają wybrane przykłady wyników prac testowych z zakresu transformacji polskich danych przestrzennych do modeli INSPIRE. Otrzymane zapisy danych w języku GML (rys. 2.19) nie są jeszcze w pełni zgodne z wymaganiami specyfikacji, ale są jak najbardziej jak to było możliwe, zbliżone do zapisów całkowicie spełniających te wymagania.

2.8. Transformacja zbiorów danych w językach aplikacyjnych XML

Opisywane prace testowe z zakresu transformacji wykazały, że dostępne obecnie oprogramowanie nie pozwala na pełną transformację polskich danych do schematów INSPIRE. Konieczne jest w tym przypadku zastosowanie technologii opartej o XSLT (*XSL Transformations, Extensible Stylesheet Language Transformations*). Technologia ta pozwala na transformację dowolnego zapisu w języku XML do innego zapisu XML opartego na zupełnie innym schemacie XSD (rys. 2.23). Reguły transformacji zapisane są w pliku XML o nazwie arkusz XSLT, a sam proces transformacji jest realizowany przez oprogramowanie nazywane procesorem XSLT (Mangano, 2007). Do najbardziej popularnych procesorów XSLT należą: Saxon (Java oraz .NET), AltovaXML 2010 (linia komend, COM, Java, .NET), TransforMiiX (C++), Xalan (Java, C++), biblioteka libxslt/xsltproc (C) i funkcje XSLT w PHP 5 (Wikipedia, 2010).



Rys. 2.23. Prosty przykład procesu transformacji danych zapisanych w aplikacjach języka XML za pomocą procesora XSLT, dla którego reguły transformacji są zapisane w arkuszu XSLT – jest to również zapis w schemacie XML (W3C, 1999; Wikipedia, 2010)

Przykładem zastosowania technologii XSLT do danych przestrzennych jest zobrazowanie danych zapisanych w języku GML za pomocą graficznego obrazu wektorowego zapisanego w języku SVG (*Scalable Vector Graphics*). Takie zobrazowanie jest cennym rozszerzeniem standardowej usługi WMS, w której najczęściej tworzony obraz ma zapis rastrowy. Obszerny opis takiego zastosowania zawiera praca Tennakoona (2003), a także inne publikacje (Galdos, 2002). Inne zastosowanie XSLT do GML to transformacja danych OSM (*Open Street Map*) do GML (OSM, 2009).

W zakresie zastosowania technologii XSLT do danych INSPIRE prace są dopiero na wczesnym etapie, m.in. w ramach projektu INSPIRE-FOSS (*Free and Open Source Software*). Na stronach tego projektu (INSPIRE-FOSS, 2010) umieszczane są aktualnie opracowywane pliki XSLT dla zastosowań w usłudze WFS (*Web Feature Service*), dotyczące danych z tematów: adresy, działki ewidencyjne i nazwy geograficzne. Jednak praktyczne zastosowanie XSLT do transformacji danych do schematów INSPIRE wymaga jeszcze wiele pracy, a w przypadku danych polskich zadanie to spoczywa na krajowym środowisku osób zainteresowanych tymi zagadnieniami.

2.9. Podsumowanie

Transformacja polskich danych przestrzennych do modeli INSPIRE jest jednym z podstawowych zadań, jakie stoi przed polskim środowiskiem osób zajmujących się implementacją dyrektywy INSPIRE. Jest to zadanie trudne, ponieważ tego rodzaju prace, w tak szerokim zakresie i jednocześnie z tak wysoko postawionymi wymaganiami, nie były w Polsce dotychczas realizowane.

Przedstawione tu prace studialne i testowe, dotyczące aspektu technologicznego tej transformacji, pozwalają na sformułowanie kilku wniosków i uwag dotyczących metodyki realizacji tego zadania. Prace testowe dotyczyły małych próbek danych przestrzennych, głównie z zakresu tematów pierwszej grupy, jednak uzyskane wyniki dają się ekstrapolować do skali znacznie większej, w której będą transformowane duże zasoby obejmujące terytorium całego kraju. Główne wnioski i uwagi można sformułować następująco:

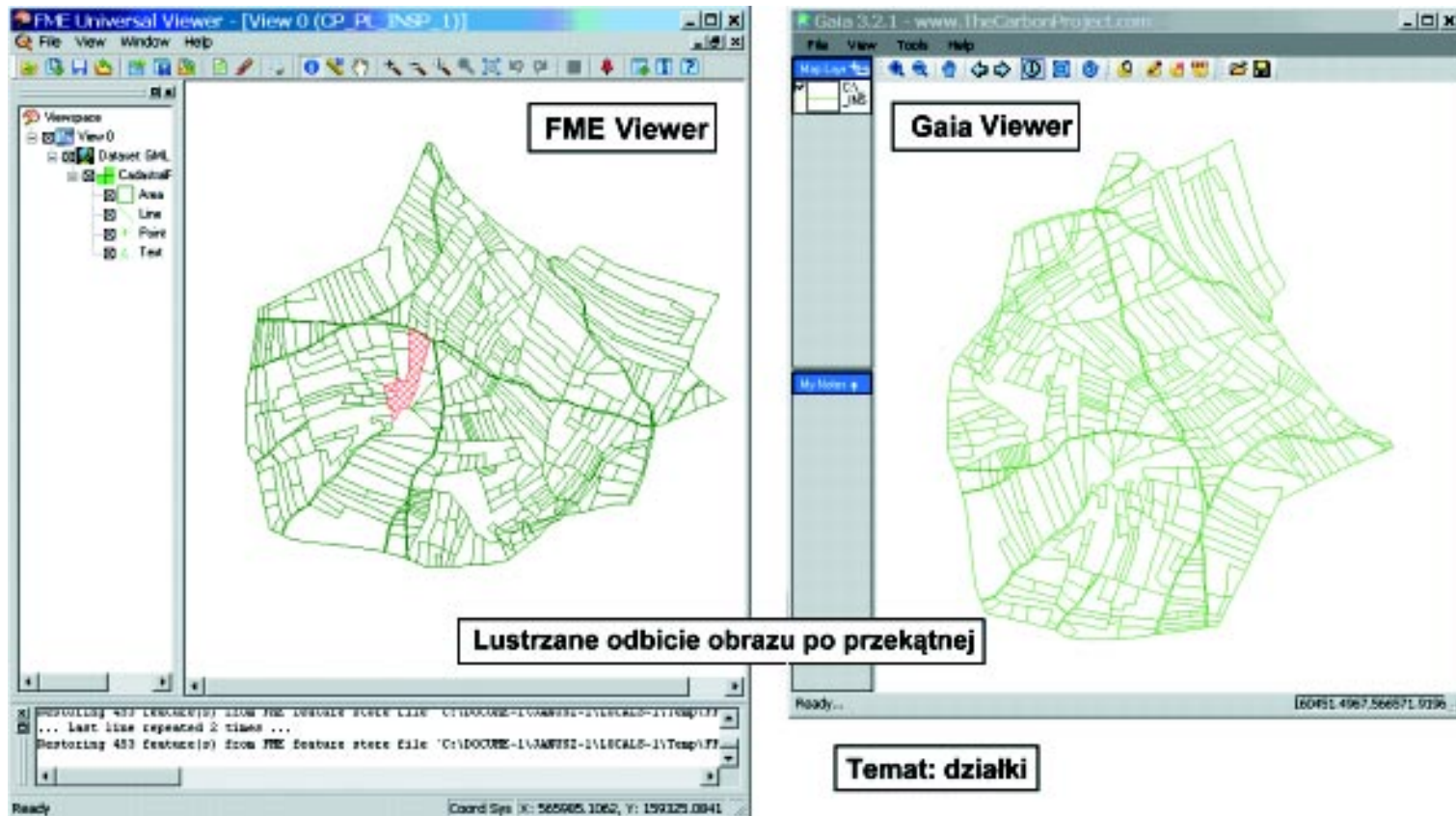
1. Na obecnym etapie transformacja polskich danych powinna być wykonana w trybie *off-line*, ponieważ różnice pomiędzy modelami danych źródłowych i wynikowych są bardzo duże, a często dla utworzenia jednego zbioru wynikowego potrzebne są dane z kilku zbiorów źródłowych.
2. Szczegółowego rozważenia wymaga problem utrzymywania danych w ich obecnej postaci, w sytuacji istnienia zbiorów spełniających wymagania specyfikacji INSPIRE i zawierających wszystkie dane niezbędne dla obecnych potrzeb krajowych. Równoległe utrzymywanie dwóch różnych kategorii zasobów w wielu dziedzinach jest bardzo kosztowne i prawdopodobnie niepotrzebne.
3. Proponuje się podzielenie procesu transformacji w trybie *off-line* na trzy etapy. W pierwszym zostanie ujednoczona forma zapisu do języka GML, kodowania UTF-8 i jednego wybranego układu odniesienia np. EPSG:2180 (Poland CS 92). W drugim etapie trzeba będzie ujednoczyć wszystkie pozostałe elementy danych, w tym format zapisu liczb, identyfikatory wyróżnień i elementy niezbędne do utworzenia powiązań między wyróżnieniami. Trzeci etap to migracja danych źródłowych do zbiorów w pełni zgodnych ze specyfikacjami INSPIRE.
4. Trzeci etap wymaga zastosowania technologii XSLT. To zagadnienie nie jest jeszcze opracowane dostatecznie poprawnie. Wymaga ono dalszych prac studialnych i testowych, z uwzględnieniem znacznych różnic jakie występują pomiędzy poszczególnymi tematami INSPIRE, także tymi, dla których specyfikacje są dopiero na etapie opracowywania.

Literatura

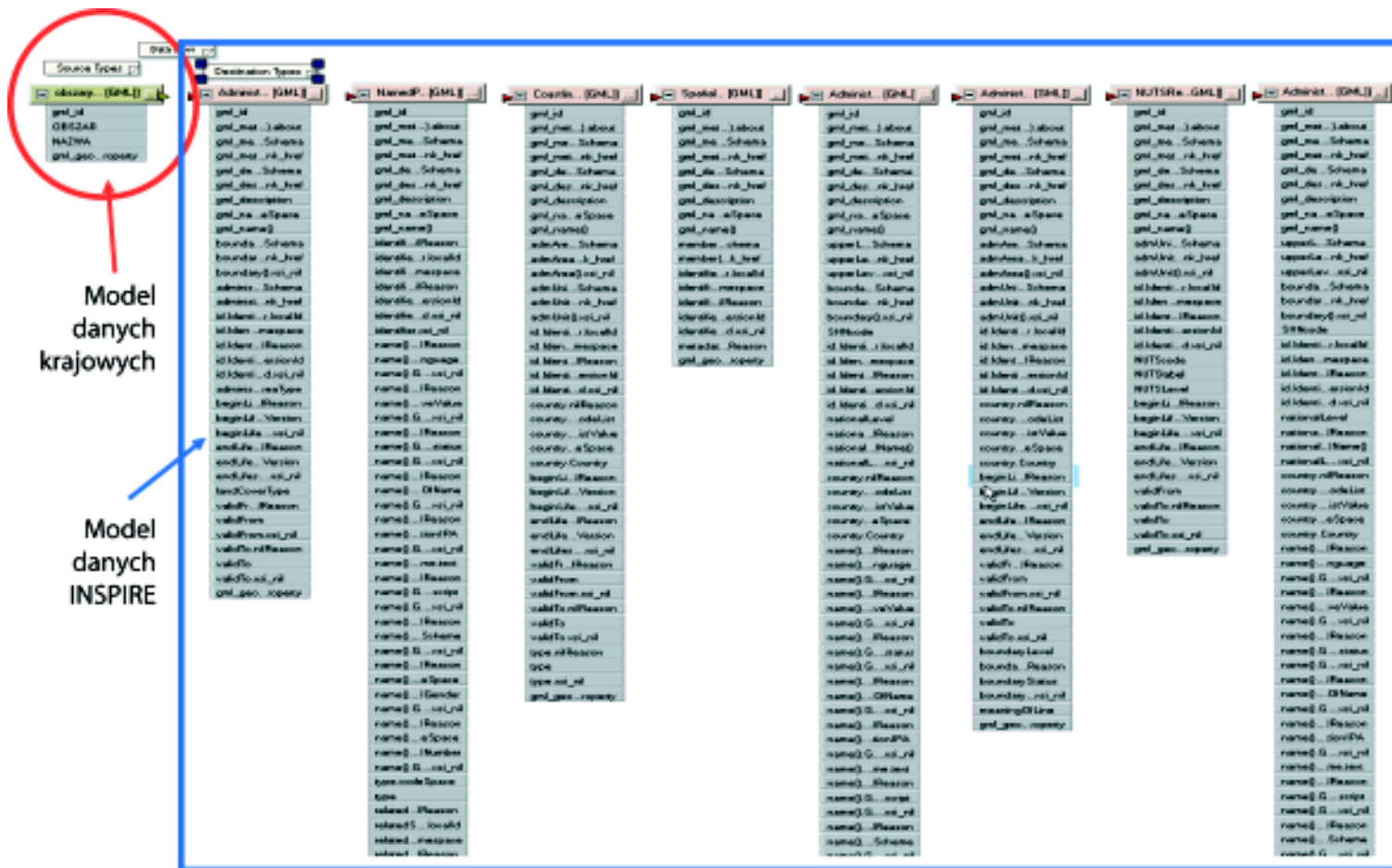
- Bray T., Paoli J., Sperberg-Mcqueen C. M., 1998: Extensible Markup Language (XML) 1.0, W3C Recommendation. URL: <http://www.w3.org/TR/1998/REC-xml-19980210>
- Carlson D., 2001: Modeling XML Applications with UML: Practical e-Business Applications. Addison-Wesley, Boston.
- ESRI, 2010: Esri Technology for INSPIRE.
URL: <http://www.esri.com/library/whitepapers/pdfs/esri-technology-and-inspire.pdf>
- Galdos Systems, 2002: GML XSLT Style Extensions.
URL: <http://sourceforge.net/projects/gml-xslt-ext/files/>
- Herring J., Kottman C., 1997: Features and Coverages for the Layman. OpenGIS Newsletter, 2,4: 6-7.
- IGiK, 2009: Prace nad identyfikacją zbiorów i usług danych przestrzennych dla I i II grupy tematycznej INSPIRE. Archiwum GUGiK.

- INSPIRE, 2007: Methodology for the Development of data Specifications (D 2.6, Version 2.0). Report, INSPIRE Doc. Repository.
- INSPIRE, 2008a: Definition of Annex Themes and Scope (D 2.3, Version 3.0). Report, INSPIRE Doc. Repository.
- INSPIRE, 2008b: Generic Conceptual Model (D 2.5, Version 3.1). Report, INSPIRE Doc. Repository.
- INSPIRE, 2008c: Guidelines for the encoding of spatial data (D 2.7, Version 3.0). Report, INSPIRE Doc. Repository.
- INSPIRE, 2008d: INSPIRE Feature Concept Dictionary. Register, INSPIRE Doc. Repository.
- INSPIRE, 2008e: INSPIRE Glossary. INSPIRE Doc. Repository.
- INSPIRE, 2008f: Methodology for the development of data specifications: baseline version (D 2.6, Version 3.0). Report, INSPIRE Doc. Repository.
- INSPIRE, 2008g: Methodology for the development of data specifications: List of Abbreviations. Report, INSPIRE Doc. Repository.
- INSPIRE, 2009a: Draft Implementing Rules for INSPIRE Transformation Services (Version 3.0). INSPIRE Doc. Repository.
- INSPIRE, 2009b: INSPIRE Draft Transformation Implementing Rule (Version 2.0). INSPIRE Doc. Repository.
- INSPIRE, 2010a: Draft Technical Guidance for INSPIRE Coordinate Transformation Services. INSPIRE Doc. Repository.
- INSPIRE, 2010b: GML Application Schemas (April 2010). INSPIRE Doc. Repository.
- INSPIRE, 2010c: INSPIRE Code List Dictionaries (April 2010). INSPIRE Doc. Repository.
- INSPIRE, 2010d: INSPIRE Consolidated UML Model (April 2010). INSPIRE Doc. Repository.
- INSPIRE, 2010e: INSPIRE Data Specification on Administrative Units – Guidelines v3.0.1. INSPIRE Doc. Repository.
- INSPIRE, 2010f: INSPIRE Data Specification on Cadastral Parcels – Guidelines v 3.0.1. INSPIRE Doc. Repository.
- INSPIRE, 2010g: INSPIRE Data Specification on Geographical Names – Guidelines v 3.0.1. INSPIRE Doc. Repository.
- INSPIRE, 2010h: INSPIRE Data Specification on Hydrography – Guidelines v 3.0.1. INSPIRE Doc. Repository.
- INSPIRE, 2010i: INSPIRE Data Specification on Protected Sites – Guidelines v 3.1.0. INSPIRE Doc. Repository.
- INSPIRE, 2010j: INSPIRE Data Specification on Transport Networks – Guidelines v 3.1. INSPIRE Doc. Repository.
- INSPIRE, 2010k: INSPIRE Data Specifications on Addresses – Guidelines v 3.0.1. INSPIRE Doc. Repository.
- INSPIRE, 2010l: INSPIRE draft Domain Model. Technical Report, INSPIRE Doc. Repository.
- INSPIRE, 2010m: INSPIRE Specification on Coordinate Reference Systems – Guidelines v 3.1. INSPIRE Doc. Repository.
- INSPIRE, 2010n: INSPIRE Specification on Geographical Grid Systems – Guidelines v 3.0.1. INSPIRE Doc. Repository.
- INSPIRE-FOSS, 2010: Transformation, storage and web-based delivery for geodata based on European INSPIRE standards using Free and Open Source Software. URL: <http://code.google.com/p/inspire-foss>
- Intergraph, 2010: Case study: EU SD I Research Projects – Intergraph – Taking a Lead in EU Programmes. URL: <http://www.intergraph.com/assets/pdf/INSPIREprojectsScreen.pdf>
- Lake R., 1999: Introduction to GML – Geography Markup Language, Galdos Systems Inc. Arch. URL: <http://www.focalpoint.org/galdos/GMLIntroduction.html>
- Larman C., 2001: Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and the Unified Process (2nd Edition). Prentice Hall, New Jersey.
- Lehto. L., Sarjakoski T., 2004: Schema translations by XSLT for GML encoded geospatial data in heterogeneous Webservice environment. Proceedings of the XXth ISPRS Congress, July 2004, Istanbul, Turkey, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXV (B4/IV): 177-182.
- Mangano S., 2007: XSLT – receptury. HELION, Gliwice.

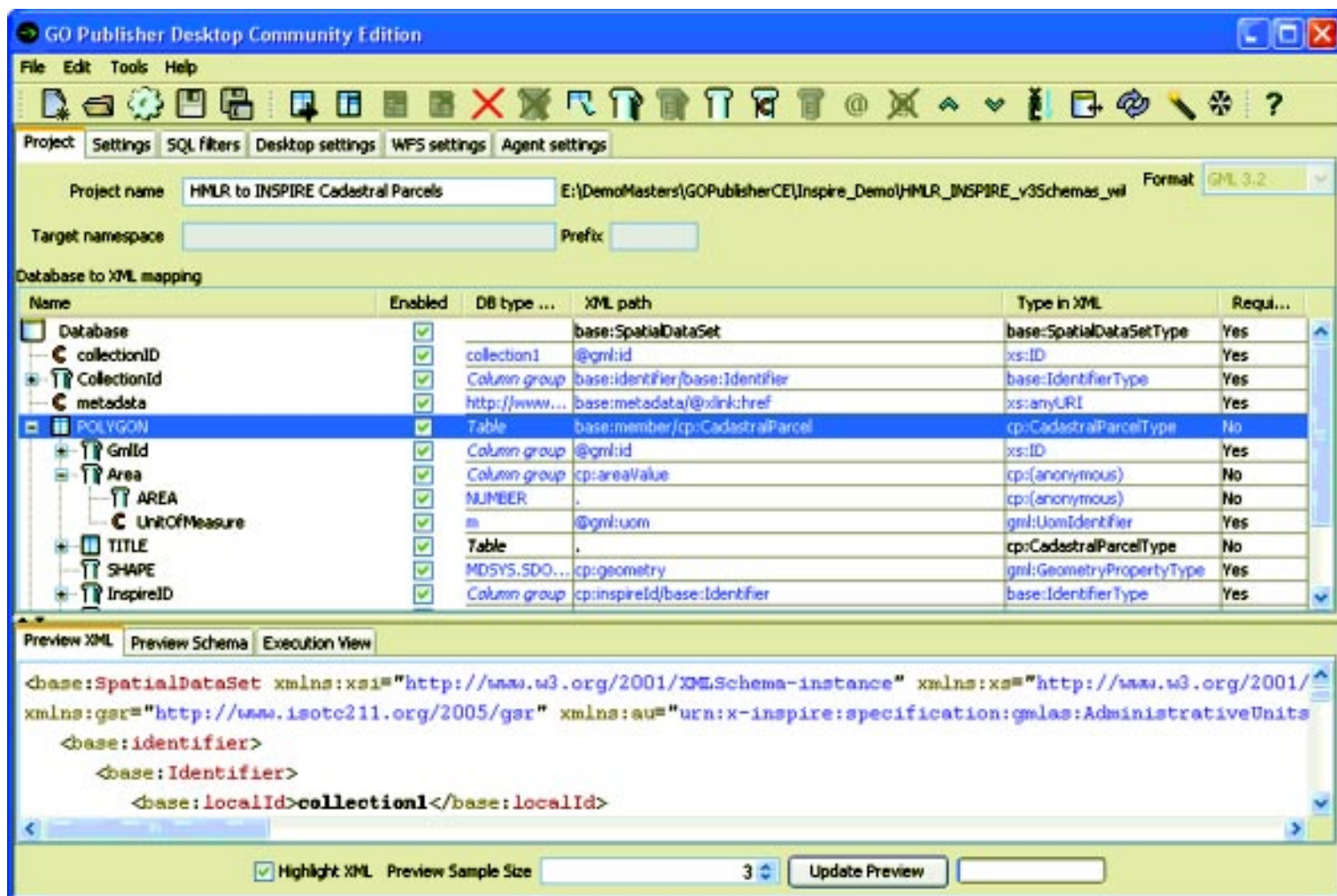
- Mark D. M., Skupin A., Smith B., 2001: Features, Objects, and other Things: Ontological Distinctions in the Geographic Domain. Spatial Information Theory, Proceedings of COSIT 2001, Springer.
URL: <http://wings.buffalo.edu/philosophy/faculty/smith/articles/COSIT01MSS.pdf>
- Mercer D., 2000: XML – kurs podstawowy. Wyd. Edition 2000, Kraków.
- Michalak J., 2003: Podstawy metodyczne i technologiczne infrastruktur geoinformacyjnych. *Roczniki Geomatyki*, t. 1, z. 2, monografia, 140 str. PTIP, Warszawa.
- Michalak J., 2005: Dziedziczne modele pojęciowe dotyczące informacji przestrzennej na przykładzie geologii. *Roczniki Geomatyki*, t. 3, z. 3: 135-146. PTIP, Warszawa.
- Michalak J., 2008: Języki pochodne od GML i z nim powiązane. *Roczniki Geomatyki*, t. 6, z. 5: 75-84. PTIP, Warszawa.
- Michalak J., 2010: Metodyka i technologia budowy geoserwera tematycznego jako komponentu INSPIRE. *Roczniki Geomatyki*, t. 8, z. 3, monografia. PTIP, Warszawa.
- OGC, 2007: OpenGIS Geography Markup Language (GML) Encoding Standard. Ver. 3.2.1, Doc. Nr: 07-036, Ed.: Clemens Portele. OGC Repository, Wayland.
- OSM (Open Street Map Community), 2009: Converting OSM to GML.
URL: http://wiki.openstreetmap.org/wiki/Converting_OSM_to_GML
- Safe Software, 2010: FME INSPIRE Protected Areas WFS Prototype.
URL: http://fmepedia.com/index.php/INSPIRE_ProtectedAreas_Demo
- Schneider M., 2010: Implementing INSPIRE Data Themes with deegree 3.
URL: <http://wiki.deegree.org/deegreeWiki/deegree3/InspireSupport>
- Skogan D., 1999: UML as a Schema Language for XML based Data Interchange.
URL: <http://www.cs.colostate.edu/UML99/skoganUMLpaper.pdf>
- Smith B., 2001: Fiat Objects. Topoi.
URL: <http://wings.buffalo.edu/philosophy/faculty/smith/articles/fatobjects.pdf>
- Snowflake Software, 2010: Our SolutionL Get GO'ing with GO Publisher Desktop Community Edition.
URL: <http://www.snowflakesoftware.com/markets/inspire/solution.htm>
- Tennakoon W. T., 2003: Visualization of GML data using XSLT.
URL: http://www.itc.nl/library/papers_2003/msc/gim/tennakoon.pdf
- W3C, 1999: XSL Transformations (XSLT) Version 1.0. URL: <http://www.w3.org/TR/xslt>
- Wikipedia, 2010: XSL Transformations. URL: http://pl.wikipedia.org/wiki/XSL_Transformations.



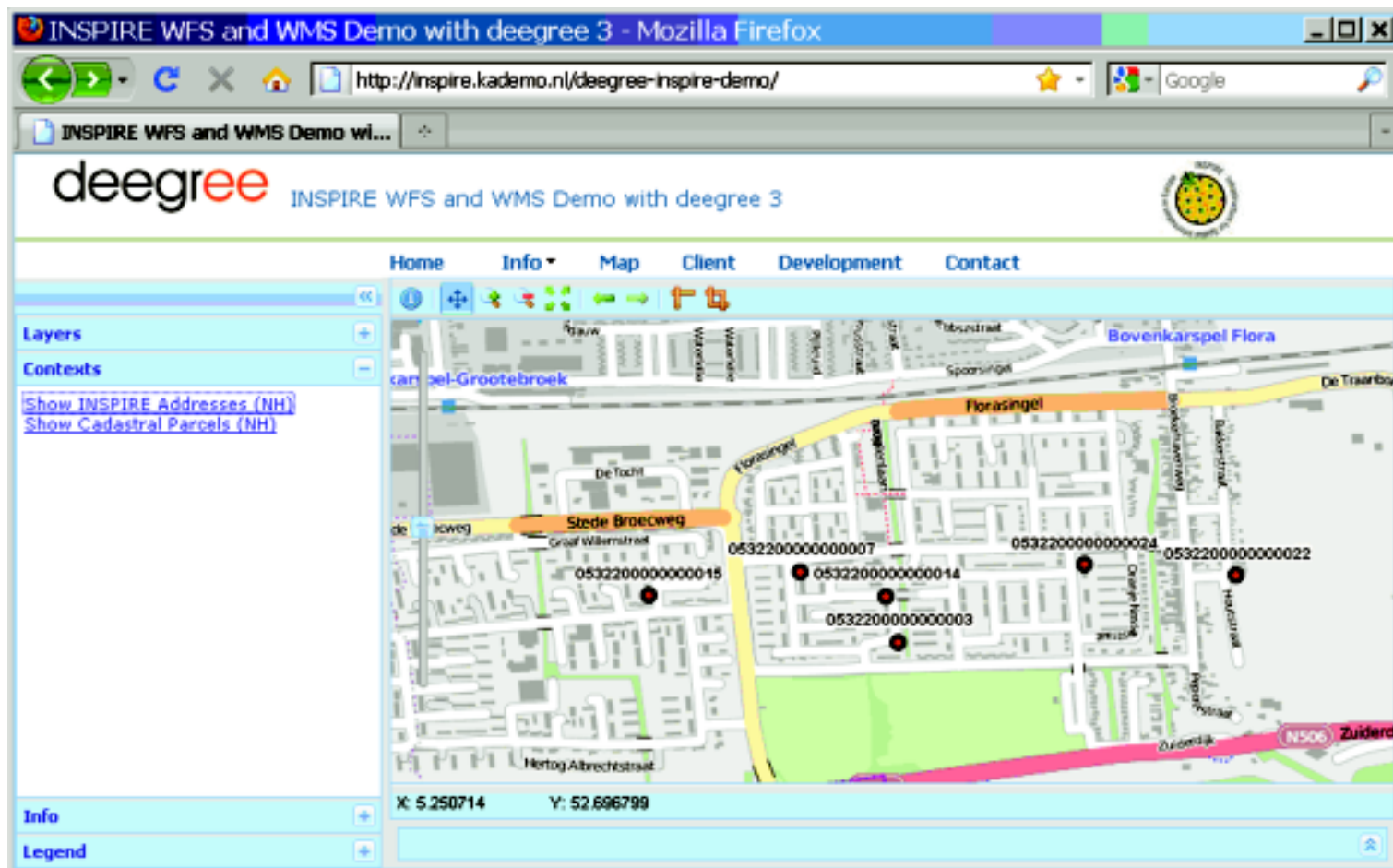
Rys. 2.9. *Axis order* (kolejność współrzędnych) – znany problem Języka GML. W różnych wersjach tego języka, a także w różnych sposobach zapisu w ramach jednej wersji kolejność współrzędnych może być różna dla różnych układów odniesienia i w konsekwencji różne programy interpretują to różnie. Przykład przedstawia zobrazowanie tego samego pliku przy pomocy dwóch zaawansowanych przeglądark GML



Rys. 2.10. Porównanie liczby elementów i ich atrybutów pomiędzy modelem danych polskich (czerwone koło) i modelem INSPIRE (niebieski prostokąt) dla tematu *jednostki administracyjne*. W tym przypadku dane polskie to dane wyprowadzone z bazy danych Państwowego Rejestru Granic Administracyjnych w formacie SHP. Fragment okna programu FME Workbench

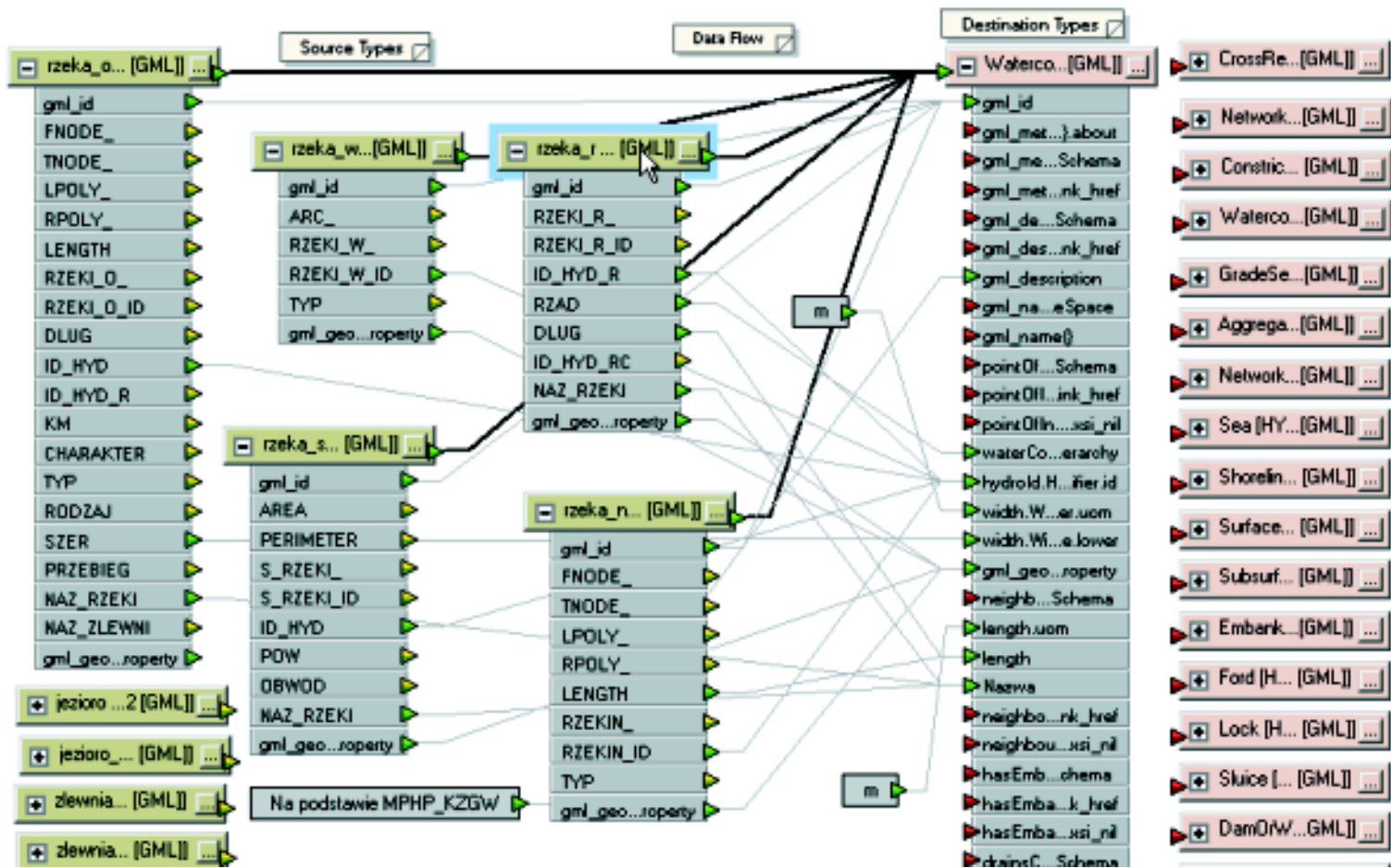


Rys. 2.11. Okno programu GO Publisher (Desktop Community Edition) – jednego z trzech najbardziej zaawansowanych systemów do transformacji danych pomiędzy różnymi modelami (Snowflake, 2010). Przykład dotyczy mapowania danych tematu *działki ewidencyjne*

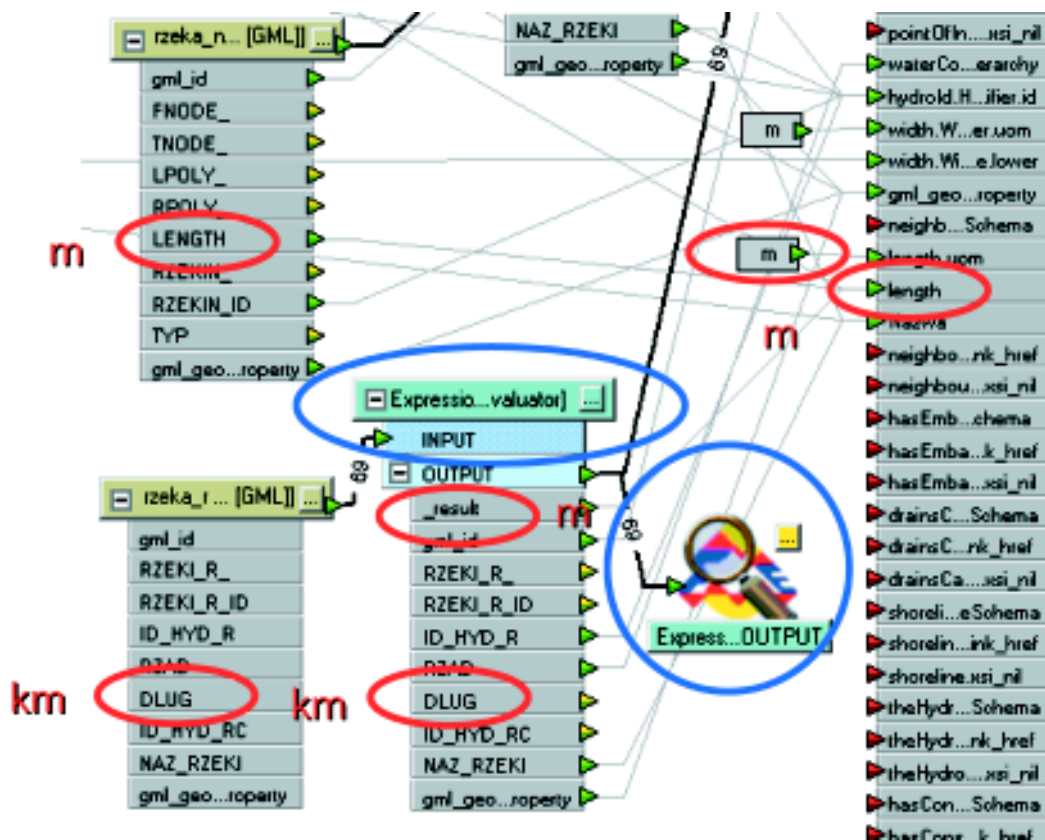


Rys.2.12. Okno przeglądarki WWW Firefox przedstawiające przykład zastosowania systemu Deegree 3 w geoserwerze INSPIRE w zakresie tematów adresy i działki ewidencyjne (Schneider, 2010)





Rys. 2.16. Przykład mapowania elementów danych przestrzennych i ich atrybutów z modelu polskiego do modelu INSPIRE tematu hydrografia. Dane z kilku plików źródłowych są łączone w jeden plik wynikowy. Fragment okna programu FME Workbench



Rys. 2.17. Przykład zastosowania transformera do przekształcania atrybutu liczbowego (czerwone elipsy) z modelu polskiego do modelu INSPIRE tematu hydrografia. W modelu polskim długość odcinka rzeki jest podana w kilometrach a w modelu INSPIRE ta wielkość powinna być podana w metrach.

Transformer *Expression evaluator* (niebieska elipsa) pozwala przeliczyć dowolne wielkości liczbowe w trakcie transformacji i wynik przeliczenia może być obserwowany za pomocą programu FME Viewer połączonego z elementem Expression Output (niebieskie koło).

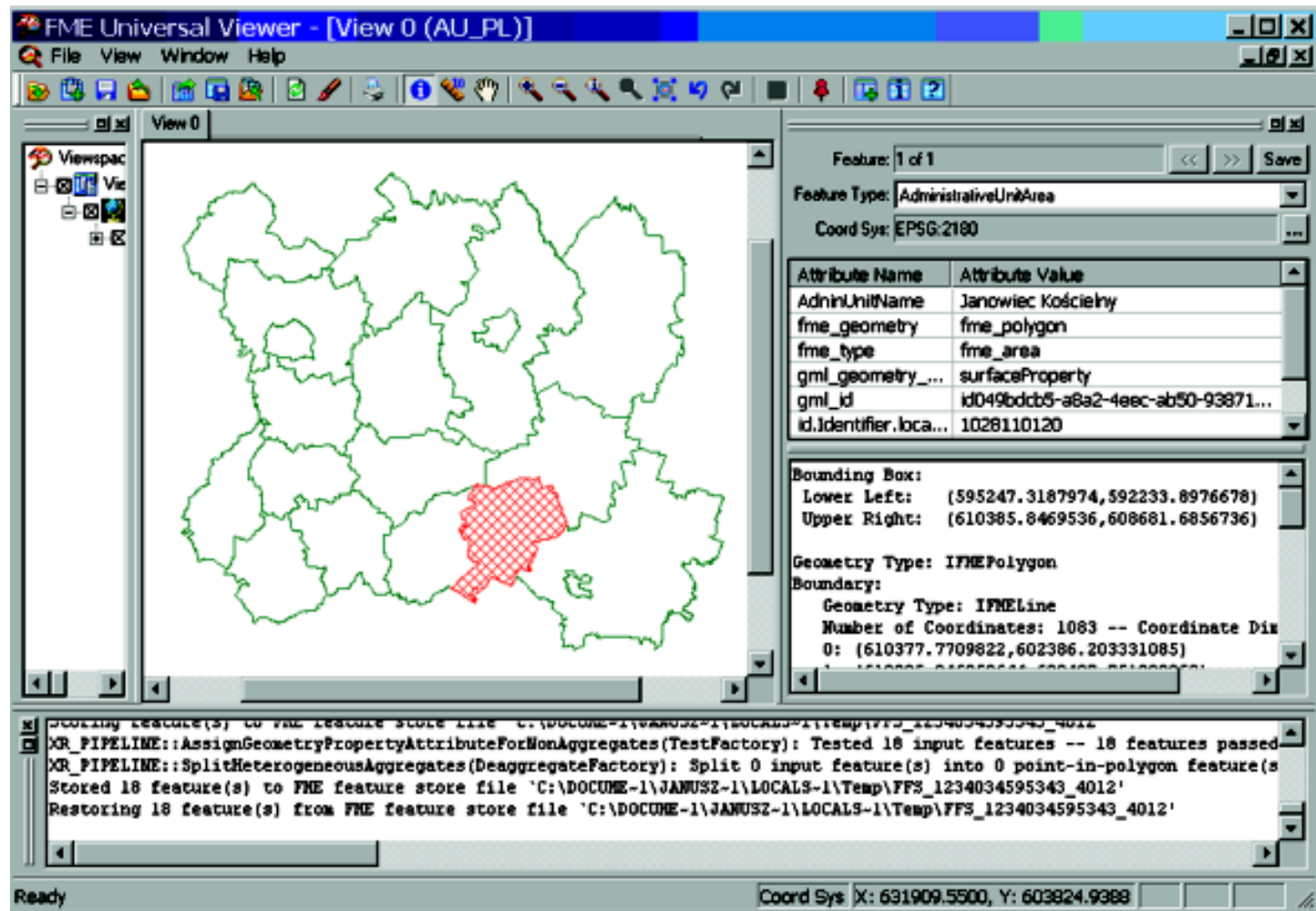
Fragment okna programu FME Workbench

```

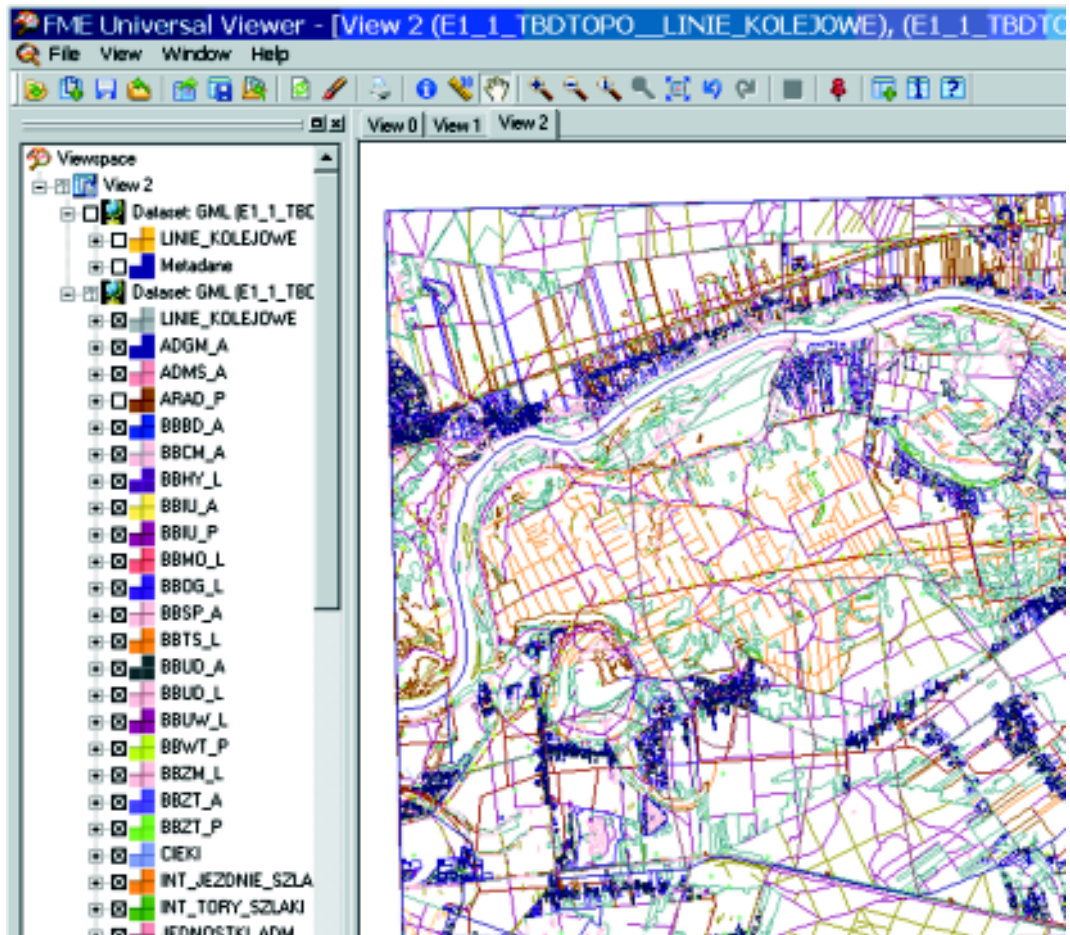
<?xml version="1.0" encoding="UTF-8" ?>
- <gml:FeatureCollection xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:AU="urn:x-INSPIRE:specification:gmlas-
  v31:AdministrativeUnits:2.0" xsi:schemaLocation="urn:x-INSPIRE:specification:gmlas-v31:AdministrativeUnits:2.0
  AdministrativeUnitsPL.xsd">
- <gml:boundedBy>
- <gml:Envelope srsName="EPSG:2180" srsDimension="2">
  <gml:lowerCorner>586141.802749292 560295.401295859</gml:lowerCorner>
  <gml:upperCorner>648008.588742649 630714.9344332</gml:upperCorner>
</gml:Envelope>
</gml:boundedBy>
- <gml:featureMember>
- <AU:AdministrativeUnitArea gml:id="Id9c73c893-5372-4f4a-874b-015ad3b21e8a">
  <gml:description>National Boundary Register</gml:description>
- <AU:id>
- <AU:Identifier>
  <AU:localId>1014020520</AU:localId>
  <AU:namespace>PL_AU</AU:namespace>
</AU:Identifier>
</AU:id>
<AU:AdminUnitName>Grudusk</AU:AdminUnitName>
<AU:gml_geometry_property>surfaceProperty</AU:gml_geometry_property>
- <gml:surfaceProperty>
- <gml:Surface srsName="EPSG:2180" srsDimension="2">
  - <gml:patches>
  - <gml:PolygonPatch>
  - <gml:exterior>
  - <gml:LinearRing>
    <gml:posList>603914.604353963 574127.633106319 603915.744706518
    574167.051283229 603912.461626365 574172.878816404 603891.490139371
    574217.478626308 603887.267208806 574229.419945949 603888.235978339
    574235.925422784 603902.504658833 574295.047348958 603905.671969563
    574301.59499188 603918.703412303 574315.107130065 603921.497278209
    574322.133954125 603926.58361927 574350.913268645 603943.439089901
    574378.599212735 603985.403434486 574428.454295527 604009.490723899
    574471.132248018 604035.431756039 574497.14306878 604043.239517677
    574496.820419029 604122.686668919 574521.34451916 604133.828534262
    574523.384904349 604243.684628414 574532.048855286 604253.679542416
    574534.707654421 604264.81015978 574540.727332533 604279.521093807
    574549.587041041 604283.119368446 574552.506778309 604291.237380346
    574553.243630386 604296.227985346 574553.202467328 604300.717423987

```

Rys. 2.19. Fragment początkowy zapisu polskich danych przestrzennych z Państwowego Rejestru Granic Administracyjnych w schemacie INSPIRE dla tematu jednostki administracyjne. Zapis ten nie jest w pełni zgodny ze specyfikacją dla tego tematu, ponieważ nie było możliwości transformacji wszystkich wymaganych elementów i atrybutów



Rys. 2.20. Okno programu FME Viewer przedstawiające zobrazowanie próbki danych z Państwowego Rejestru Granic Administracyjnych zapisanych w języku GML w sposób częściowo zgodny ze schematem aplikacyjnym INSPIRE (rys. 3.19)



Rys. 2.21. Dane TBD (Topograficznej Bazy Danych) zapisane w schemacie aplikacyjnym TBDGML wersja 1.33 przedstawione graficznie za pomocą programu FME Viewer

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <Plik_TBD xmlns="http://www.gugik.gov.pl/TBD" xmlns:gml="http://www.opengis.net/gml"
  xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="
  http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
  http://www.gugik.gov.pl/TBDGML_ver1_33.xsd">
3   <Metadane>
24   <Dane>
25     <ULICE>
26       <ID>7193</ID>
27       <NAZWA_CZ2>Rolna</NAZWA_CZ2>
28       <ID_MIEJSCOWOSC>0918123</ID_MIEJSCOWOSC>
29       <X_KAT_DOK_ATRYB>1</X_KAT_DOK_ATRYB>
30       <X_AKTUALNOSC_ATRYB>2003-12-01</X_AKTUALNOSC_ATRYB>
31     </ULICE>
32     <ULICE>
33       <ID>73</ID>
34       <NAZWA_CZ2>Boksterska</NAZWA_CZ2>
35       <ID_MIEJSCOWOSC>0918123</ID_MIEJSCOWOSC>
36       <X_KAT_DOK_ATRYB>1</X_KAT_DOK_ATRYB>
37       <X_AKTUALNOSC_ATRYB>2003-12-01</X_AKTUALNOSC_ATRYB>
38     </ULICE>
39     <ULICE>
47     </ULICE>
```

Rys. 2.22. Początkowy fragment jednego z wielu plików zapisu w schemacie TBDGML wersji 1.33 zawierający nazwy ulic. Struktura i treść tych plików znacznie odbiega od wymagań specyfikacji danych INSPIRE, co powoduje duże problemy przy ich transformacji. Okno programu XML Spy do analizy i edycji zapisów XML i ich schematów (XSD)

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 methodological 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