



**POLSKIE
TOWARZYSTWO
INFORMACJI
PRZESTRZENNEJ**

ROCZNIKI 2010

GEOMATYKI

Modelowanie danych przestrzennych

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

Elżbieta Bielecka

3. ZASADY OCENY JAKOŚCI DANYCH PRZESTRZENNYCH ORAZ ICH ZASTOSOWANIE DO OCENY JAKOŚCI DANYCH GROMADZONYCH W TBD

3.1. Wprowadzenie

Wobec powszechnego dostępu do danych przestrzennych znajomość ich jakości staje się zagadnieniem kluczowym zarówno dla użytkowników danych, jak też instytucji je udostępniających. Jakość samą w sobie trudno jest zdefiniować, pod tym pojęciem kryje się bowiem wszystko to, co ma związek z pewnymi cechami produktu (wyrobu) lub usługi, mającymi wpływ na postrzeganie go przez klienta. Uważa się, że pierwszy pojęcie jakości (gr. *poiotēs*) zdefiniował Platon jako „pewien stopień doskonałości”. Było to wówczas pojęcie filozoficzne i jako takie pozostało do czasów współczesnych. Jakość zatem, jest pojęciem subiektywnym, postrzeganym przez pryzmat własnej oceny, inaczej przez użytkownika, producenta, czy dystrybutora danych. W literaturze znaleźć można wiele definicji jakości, w tym również w odniesieniu do danych przestrzennych. Z reguły uważa się, że jakość „producenta” to zgodność wyrobu ze specyfikacjami technicznymi, natomiast jakość „użytkownika” to możliwość zaspokojenia jego wymagań przez określony produkt.

Celem warsztatów „Modelowanie danych przestrzennych”, w zakresie oceny jakości danych przestrzennych, jest zapoznanie uczestników ze znormalizowanymi podstawami oceny jakości danych, zgodnie z zasadami podanymi w normach ISO serii 19100 oraz ocena zgodności danych zgromadzonych w TBD z Wytocznymi TBD, czyli ocena jakości „producenta”. W zakresie oceny danych TBD zaprezentowane zostanie przyporządkowanie elementom kontroli jakości wg ISO 19 113 (ISO, 2002) i specyfikacji technicznych INSPIRE odpowiednich zestawów automatycznych kontroli TBD *Aplikacja O*; interpretacji błędów wykrytych w wyniku kontroli oraz przedstawienie dodatkowych kontroli koniecznych do oceny jakości zbioru zgodnie z wymaganiami przepisów implementacyjnych dyrektywy INSPIRE.

3.2. Ocena jakości danych przestrzennych według norm ISO serii 19 100

Jakość danych przestrzennych jest zagadnieniem na tyle ważnym, że zajęły się nim międzynarodowe organizacje normalizacyjne. Wśród norm ISO serii 19 100, aż trzy dokumenty podejmują temat jakości danych przestrzennych¹. Należą do nich norma ISO 19113 Geogra-

¹ W normach ISO serii 19 100 używany jest termin dane geograficzne (ang. *geographical data*), natomiast w terminologii polskiej częściej używa się określenia dane przestrzenne, który jest traktowany jako odpowiednik normatywnego określenia dane geograficzne (Pachelski, 2009).

phic information – Quality principles (ISO, 2002), norma ISO 19114 Geographic information – Quality evaluation procedures (ISO, 2003a) oraz specyfikacje techniczne ISO/TS 19138 Geographic information – Data quality measures (ISO, 2006). Pierwsza z wymienionych norm ustanawia model oraz zasady zapisu jakości danych przestrzennych, w drugiej został zdefiniowany proces identyfikacji i oceny potrzeb informacyjnych w zakresie jakości oraz procedury oceny jakości danych. Natomiast dokument ISO/TS 19138 uzupełnia zakres normy ISO 19114 podając mierniki jakości danych. Wszystkie trzy dokumenty tworzą spójny zbiór wytycznych przeznaczonych do opisu jakości danych przestrzennych oraz procesu oceny jakości. Analizując rozwiązania normatywne ISO serii 19 100 w zakresie jakości Kmiecik (2004) stwierdziła, że normy ISO regulują problematykę jakości danych w sześciu aspektach:

- 1) organizacji opisu jakości,
- 2) procesu ustalania i oceny modelu jakości dla danych przestrzennych,
- 3) procedur oceny jakości danych,
- 4) metod oceny jakości danych,
- 5) miar jakości,
- 6) zapisu jakości zbiorów danych przestrzennych.

Jakość danych przestrzennych można określać w odniesieniu do zbioru danych przestrzennych, serii zbiorów danych przestrzennych lub wyodrębnionego podzbioru i zapisuje się bezpośrednio w metadanych, zgodnie z ISO 19115 (ISO, 2003b) oraz w raporcie zewnętrzny zdefiniowanym w normie ISO 19114 (ISO, 2003a).

Wymienione normy ISO z zakresu informacji geograficznej definiują jakość jako kompleksowy zespół cech i charakterystyk zbiorów danych, które wpływają na możliwość zaspokojenia przez ten zbiór wymagań użytkowników. W przypadku danych przestrzennych wymagania użytkowników są związane z wykorzystaniem ich w konkretnych aplikacjach i do konkretnych celów. Najczęściej, jakość jest określana na podstawie zestawu cech charakteryzujących zbiór danych zapisanych w metadanych. Zgodnie z normą ISO 19113:2002, która jest jednocześnie Polską Normą dostępną w języku polskim (PN-EN ISO 19113:2009), jakość danych przestrzennych charakteryzuje się określając ilościowe i przeglądowe elementy jakości danych przestrzennych. Do ilościowych elementów jakości należą: kompletność, zgodność (spójność) logiczna, dokładność położenia, dokładność czasowa i dokładność tematyczna (tab. 3.1). Natomiast do elementów przeglądowych, o charakterze jakościowym, zaliczamy: pochodzenie, przeznaczenie i wykorzystanie.

W normie ISO 19113:2005 podano następujące krótkie określenia elementów jakości (tłumaczenia terminów z uwzględnieniem *Internetowego Leksykonu Geomatycznego* Gaździcki, 2004):

- 1) kompletność (completeness) – obecność lub brak obiektów, ich atrybutów i związków,
- 2) logiczna zgodność (*logical consistency*)² – stopień zgodności z logicznymi regułami struktury danych, atrybutami i związkami, struktura danych może być pojęciowa, logiczna lub fizyczna,
- 3) dokładność położenia (*positional accuracy*) – dokładność pozycyjna czyli dokładność położenia obiektów,
- 4) dokładność czasowa (*temporal accuracy*) – dokładność atrybutów czasowych i czasowych związków obiektów,

² Angielski termin *logical consistency* w PN-EN ISO 19113 został przetłumaczony jako *spójność logiczna*. Termin ten będzie używany konsekwentnie w dalszej części tekstu.

5) dokładność tematyczna (*thematic accuracy*) – dokładność ilościowych atrybutów, a także poprawność nieilościowych atrybutów oraz klasyfikacji atrybutów i ich związków.

Kompletność oznacza zatem stosunek danych zgromadzonych w systemie do danych, które powinny być zgromadzone i dotyczy zarówno obszaru, który dane pokrywają, jak i atrybutów opisujących obiekty przestrzenne. Brak obiektów lub ich atrybutów określa się jako pominięcie lub niedmiar. Nadmiar danych spowodowany np. podwójną digitalizacją nazywany jest przeładowaniem. Kompletność jest najczęściej wyrażana w procentach.

Zgodność (tłumaczona również jako spójność) dotyczy wzajemnej zgodności między danymi zgromadzonymi w bazie a wartościami zdefiniowanymi w modelu danych i w specyfikacjach technicznych. Zgodność (spójność) można wyrażać w procentach lub podawać liczbę elementów niespełniających zadanych warunków.

Dokładność oznacza bliskość wartości odniesienia, niekiedy przyjmowanej za prawdziwą. W odniesieniu do danych przestrzennych mówimy o dokładności położenia (lokalizacji) obiektu, o dokładności atrybutów opisowych (dokładności tematycznej) oraz o dokładności czasu.

Tabela 3.1. Ilościowe elementy jakości danych przestrzennych wg ISO 19 113:2002

Elementy i podelementy jakości danych przestrzennych	Charakterystyka podelementów jakości danych przestrzennych
Kompletność	
– nadmiar	dane nadmiarowe w zbiorze danych
– pominięcie	dane brakujące w zbiorze danych
Spójność logiczna	
– pojęciowa	zgodność ze schematem pojęciowym
– dziedzinowa	przynależność wartości do ich dziedzin
– topologiczna	poprawność zakodowanych charakterystyk topologicznych zbioru danych
– formatu	stopień zgodności danych z fizyczną strukturą zbioru danych
Dokładność położenia	
– dokładność bezwzględna lub zewnętrzna	bliskość wartości współrzędnych do wartości przyjmowanych za prawdziwe
– dokładność względna lub wewnętrzna	bliskość wzajemnych położen obiektów danego zbioru danych do ich wzajemnych położen przyjmowanych za prawdziwe
– dokładność położenia danych gridowych	bliskość wartości określających położenie danych gridowych do wartości przyjmowanych za prawdziwe
Dokładność czasowa	
– dokładność pomiaru czasu	poprawność odniesień czasowych (dotyczy błędu w pomiarze czasu)
– spójność (zgodność) czasowa	poprawność uporządkowania wydarzeń w czasie
– ważność czasowa	ważność danych pod względem czasu (aktualność)
Dokładność tematyczna	
– poprawność klasyfikacji	porównanie klas przyporządkowanych obiektom lub ich atrybutom z przestrzenią rozważań, np. z rzeczywistością w terenie lub z referencyjnym zbiorem danych
– poprawność atrybutów nieilościowych	zgodnie z nazwą podelementu
– dokładność atrybutów ilościowych	zgodnie z nazwą podelementu

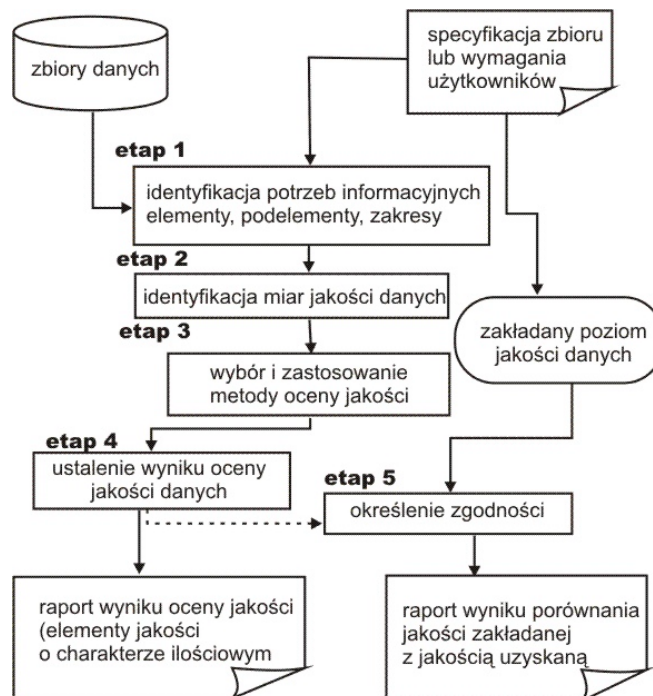
Pochodzenie, przeznaczenie i wykorzystanie zbioru danych są zaliczane do **nieliczbowych elementów jakości danych** (ISO:2002). Pochodzenie informuje użytkownika o sposobie i czasie pozyskania danych, materiałach źródłowych, zastosowanych metodach i technikach, uwzględnianych instrukcjach technicznych i standardach, dokonanych sprawdzeniach i ich wynikach. Przeznaczenie – w jaki sposób doszło do utworzenia zbioru danych przestrzennych oraz do jakich celów zbiór powinien być wykorzystany. W elemencie jakości dotyczącym wykorzystania podawane są dotychczasowe opisy zastosowania danych. Należy także podać przetworzenia, jakim podlegał zbiór np. zmiana formatu, zmiana układu współrzędnych. Syntetyczny opis jakości zbioru danych przestrzennych wyraża jego wiarygodność, czyli wierność z jaką zbiór ten odtwarza przestrzeń rozważań (Bielecka, 2009). Wiarygodność odnosi się do produktu końcowego.

W europejskiej infrastrukturze informacji przestrzennej INSPIRE założono, że jakość danych oceniana dla każdego zbioru danych udostępnianego za pomocą usług sieciowych (Draft Commission Regulation..., 2009) będzie zapisana w metadanych. Wymogi, o których mowa w art. 5 ust. 2 i art. 11 ust. 2 dyrektywy INSPIRE 2007/2/WE, odnoszące się do jakości i ważności danych przestrzennych spełnia się za pomocą dwóch elementów metadanych: pochodzenia i rozdzielczości przestrzennej (Rozporządzenie L 326/16, 2008). **Pochodzenie**, jak wspomniano wyżej, jest opisem historii procesu tworzenia lub ogólnej jakości zbioru danych przestrzennych. Podana tu informacja może zawierać stwierdzenie, czy zbiór danych był przedmiotem walidacji lub oceny jakości, czy stanowi wersję urzędową (jeżeli istnieje wiele wersji) i czy posiada moc prawną. **Rozdzielczość przestrzenna** odnosi się do poziomu szczegółowości zbioru danych i podawana jest jako mianownik skali mapy (dla danych wektorowych) lub terenowy wymiar piksela (dla danych siatkowych lub rastrowych).

Ponadto w celu pełniejszego scharakteryzowania zbioru danych zaleca się, aby charakteryzować go dodatkowo przez kompletność (z podaniem elementów brakujących i nadmiarowych), spójność pojęciową i topologiczną, dokładność położenia wyrażaną błędem bezwzględnym.

3.3. Procedura określania jakości danych

Procedura oceny jakości danych przestrzennych rozpoczyna się od identyfikacji potrzeb informacyjnych użytkowników, czyli określenia elementów i podelementów niezbędnych do oceny jakości danych i zakresu ich zastosowań (rys. 3.1). Polega na ustaleniu dla każdego podelementu takiej miary jakości aby najlepiej oszacować ilościowo jakość danych przestrzennych. W kolejnym kroku wybiera się metodę oceny jakości. Ustalenie wyników jakości następuje na podstawie zbiorczego zestawienia wartości miar jakości. Procedurę oceny jakości danych geograficznych kończy określenie stopnia zgodności wyników z zakładanym poziomem jakości.



Rys. 3.1. Pięciostopniowy proces oceny jakości danych geograficznych zdefiniowany w normie PN-EN-ISO 19114

3.4. Założenia kontroli jakości danych gromadzonych w TBD

Wszelkie dane przekazywane do zasobu TBD podlegają procesowi kontroli. Przeprowadzenie procesu kontroli należy do zadań ośrodków dokumentacji geodezyjnej i kartograficznej. Proces kontroli danych wspiera System Kontroli Bazy Danych Topograficznych (SKBDT), który jest elementem Bazy Danych Topograficznych. Głównym zadaniem SKBDT jest zautomatyzowanie i zobjektywizowanie procesu kontroli danych. Kontrola ta dotyczy zarówno poprawności technologicznej, tj. sposobu zapisu danych, parametrów technicznych (np. topologii), zgodności ze standardami wymiany danych, jak i poprawności merytorycznej, tj. kompletności danych, spełnienia wymogów dokładnościowych i zgodności danych z rzeczywistą sytuacją terenową. Niezbędnych mechanizmów do oceny jakości danych zgromadzonych w TBD dostarczają szablony kontroli oraz *Aplikacja R* i *Aplikacja O*. Ocena jakości wykonywana jest zarówno przez realizatora (jednostkę wytwarzającą dane TBD (Koncepcja SKBDT, 2006)) jak i ośrodek dokumentacji geodezyjnej i kartograficznej.

Szablon kontroli jest plikiem konfiguracyjnym XML³ zawierającym nazwę produktu wraz z opisem specyfikacji komponentów TOPO, KARTO, ORTO, NMT i wykazem możli-

³Opublikowany na stronach GUGiK (http://www.gugik.gov.pl/gugik/w_pages/w_doc_idx.php?loc=20).

wych ich kontroli. Kontrole są pogrupowane w zestawy kontroli, grupy kontroli i kontrole elementarne. Szablon zawiera również wszystkie specyfikacje reguł kontroli, na podstawie których działają *Aplikacje R* i *O*. Za jego pomocą są wykrywane dane, które podlegają dalszej analizie. Szablon definiuje m.in., jakie klasy obiektów powinien zawierać zbiór danych, czy należą one do klas fakultatywnych czy obligatoryjnych, a także czy potrzebne są dodatkowe pliki (np. DXF, MTD). Plik szablonu (.skd) jest plikiem niejawnym, specyficznym wyłącznie dla *Aplikacji R* i *O*, a próba odczytu jego zawartości wiąże się z ingerencją w konstrukcję pliku. Wewnętrzne elementy szablonu są specyficzne wyłącznie dla jednego oprogramowania komercyjnego (rys. 3.2) oraz częściowo zapisane w sposób binarny – niemożliwy do odczytu w innym oprogramowaniu, co przeczy zadeklarowanej w *Koncepcji SKTBD* niezależności od platform sprzętowych.

Aplikacja R umożliwia realizatorowi przeprowadzenie podstawowych kontroli poprawności danych, przed przekazaniem ich do odgik. Kontrole są wykonywane przez niezależne od platformy mechanizmy WebServices, implementowane według standardów W3C. Podstawowy zakres kontroli obejmuje kontrolę struktury danych na nośniku, kontrolę plików GML (parsowanie i walidacja), badanie struktur plików TXT, MTD, DXF, ASC oraz różnego rodzaju kontrole atrybutowe.

Aplikacja R jest powszechnie dostępna i publikowana na stronie internetowej GUGiK⁴. Pozytywny wynik kontroli danych (zgodnie z *Aplikacją R*) jest warunkiem przyjęcia danych do dalszego etapu weryfikacji w *Aplikacji O*.

Aplikacja O jest wykorzystywana przez ośrodki dokumentacji geodezyjnej i kartograficznej. Pozwala na realizację kompleksowej oceny danych przekazywanych do zasobu geodezyjnego i kartograficznego z wykorzystaniem szablonu kontroli. Kontrole elementarne są wykonywane w tej aplikacji w sposób automatyczny. *Aplikacja O* podobnie jak *R* wykrywa dane, które mogą być sklasyfikowane jako błędy lub też *zdarzenia do sprawdzenia*. Te drugie podlegają dalszej analizie i interpretacji kontrolującego, czego efektem jest wynik z kontroli. W celu ułatwienia interpretacji wyników kontroli możliwe jest wczytywanie podkładu rastrowego i wyświetlanie go razem z danymi wektorowymi. Kontrola przeprowadzana przez *Aplikację O* zawiera szereg etapów począwszy od kontroli elementarnych, interpretacji błędów, po kontrolę kameralną i terenową. Po zakończeniu procesów kontroli generowany jest raport składający się z dwóch części: metadanych kontroli i raportu właściwego. Metadane zawierają informacje ogólne z przebiegu kontroli, takie jak: datę utworzenia raportu, zakres i tematykę kontroli, dane osób odpowiedzialnych za kontrolę. Natomiast raport przedstawia wyniki kontroli pogrupowane według zasad określonych w *Koncepcji SKBDT*. W przeciwieństwie do *Aplikacji R*, *Aplikacja O* nie jest powszechnie dostępna i wymaga platformy GIS konkretnego producenta (produktu Geomedia firmy Intergraph). Końcowym efektem działania aplikacji jest nadanie zbiorowi danych TBD sygnatury i włączenie go do zasobu państwowego.

SKBDT jest jednym z najpełniejszych i najspójniejszych systemów kontroli zbiorów danych przestrzennych w państwowym zasobie geodezyjno-kartograficznym i mimo podkreślanych przez wykonawców TBD wad, jego wartości nie należy deprecjonować. Nie oznacza to jednak, że SKBDT jest wystarczający oraz dostarcza wiarygodnego i wyczerpującego opisu jakości danych, zwłaszcza w kontekście norm ISO i wymagań specyfikacji implementacyjnych INSPIRE. Podstawowymi mankamentami SKBDT wg Łabaja (2009) są:

- niejawność metody kontroli, która odbywa się za zasadzie „czarnej skrzynki”,

⁴ Aplikacja SKBDT – R 1.0.0.30 (http://gugik.gov.pl/gugik/w_pages/w_doc_idx.php?loc=20).

- brak możliwości edycji i parametryzacji szablonu kontroli przez użytkownika,
- brak możliwości wglądu w istotę kontroli i tym samym możliwości rozumienia i interpretacji algorytmów kontroli,
- zależność od platformy oprogramowania.

3.5. Ocena jakości danych gromadzonych w zasobie podstawowym TBD

3.5.1. Założenia wstępne

Przykłady oceny jakości danych zgromadzonych w zasobie podstawowym TBD zostały w większości zaczerpnięte z pracy magisterskiej, wykonanej przez Monikę Wąsowską (Wąsowska, 2010) i obronionej na Wydziale Inżynierii Lądowej i Geodezji WAT w czerwcu 2010 r. Do oceny wykorzystano zbiór obejmujący dwa arkusze mapy topograficznej (N-34-140-A-c-3, N-34-140-A-c-4) z obszaru Mińska Mazowieckiego⁵, o aktualności na rok 2007 i obszarze obejmującym ok. 1160 ha.

Każdemu z pięciu ilościowych elementów oceny jakości danych, wyszczególnionych w normie ISO 19 113, przyporządkowano odpowiednie zestawy kontroli automatycznych, przeprowadzone za pomocą *Aplikacji O SKBDT* i opisane w raportach. Wyniki kontroli wymagające dodatkowego sprawdzenia i interpretacji wykrytych zdarzeń stały się wyjściem do dalszych analiz zbioru danych, wykonanych według autorskiej metody opracowanej przez Wąsowską (2010).

Ocenę jakości danych zasobu Bazy Danych Topograficznych przeprowadzono pod kątem zgodności z założeniami Wytycznych technicznych TBD, a zatem jest to typowa ocena jakości producenta. Jako zbiór referencyjny wykorzystano dane z ewidencji gruntów i budynków⁶. Do zrealizowania wymienionych kontroli oraz wizualizacji zdarzeń zarejestrowanych w raportach z SKBDT, wykorzystano program ArcGIS 9.3.

3.5.2. Metodyka oceny jakości danych zasobu podstawowego TBD i otrzymane wyniki

Ocena kompletności

Standardowa ocena kompletności zbiorów danych, wykonywana w SKBDT za pomocą zestawu kontroli *Konstrukcja produktu*, obejmuje kontrolę obecności w zbiorze wszystkich plików, poprawności ich nazw oraz struktury, gwarantującej poprawne odczytanie plików przez aplikacje zarządzające TBD. W ramach tej kontroli sprawdzana jest także kompletność występowania obiektów (w tym także duplikowanie obiektów) i atrybutów oraz relacji między nimi. W szczególności testowane są relacje pomiędzy klasami obiektów przestrzennych a listami kodowymi, różnymi listami kodowymi, listami kodowymi a klasami geometrycznymi oraz powiązania relacyjne tabel intersekcji. Umożliwia to wyszukanie błędnych powiązań relacyjnych pomiędzy obiektami i ich atrybutami np.: brakujących lub nadmiarowych wpisów w tabelach.

⁵ Zbiór *topo czesc Minsk Mazowiecki.mbd* został pozyskany z Biura Geodety Województwa Mazowieckiego.

⁶ Dane z egib zostały udostępnione przez POGDiK w Mińsku Mazowieckim.

Ponieważ kontrole standardowe nie obejmowały sprawdzenia zgodności danych z sytuacją w terenie, zostały opracowane dodatkowe testy weryfikujące. Do weryfikacji wykorzystano dane z ewidencji gruntów i budynków, udostępnione przez PODGiK w Mińsku Mazowieckim oraz ortofotomapę udostępnioną na geoportal.gov.pl. Kontroli poddano:

- 1) zbiór budynków – BBBD_A, uznając, że w tym zbiorze najłatwiej jest pominąć obiekty, ze względu na ich dużą liczbę, zmienne zagęszczenie i stosunkowo małe rozmiary;
- 2) klasę Obiekty Inne (OIPR_P), w której powinny się znaleźć, takie obiekty jak: kapliczki, pomniki czy fontanny;
- 3) klasę Budowle hydrotechniczne (BBHY_L), obejmujący m.in. jazy, śluzy i zapory;
- 4) klasę Kompleksy użytkowania terenu (KUA_A).

Wyniki oceny kompletności

Wynik *kontroli produktu* dla analizowanych zbiorów TBD był pozytywny, co oznacza, że aplikacje kontrolne nie wykryły błędów w sprawdzanych zbiorach. Niemniej jednak w wyniku dodatkowych kontroli stwierdzono, że w klasie budynków występuje zarówno niedomiar, jak i nadmiar budynków. Z jednej strony w TBD brak jest danych o 142 budynkach umieszczonych w egib (rys. 3.3), z drugiej zaś występuje tu 57 budynków, które nie są zaewidencjonowane w egib. Zmiany te dotyczą przede wszystkim budynków gospodarczych i mieszkalnych. W porównaniu z ogólną liczbą obiektów w zbiorze (liczba budynków w klasie BBBD_A: 7476), zarówno braki, jak i nadmiar obiektów nie są zbyt znaczące i wynoszą odpowiednio: pominięcie – 1,90% i nadmiar – 0,76%.

Niedomiar obiektów zaobserwowano również w klasach *Obiekty Inne* (OIPR_P) oraz *Budowle hydrotechniczne* (BBHY_L). Analizowane zbiory nie zawierały żadnych obiektów, podczas gdy takie obiekty zidentyfikowano w terenie (rys. 3.5). Na terenie Mińska Mazowieckiego są dwie kapliczki, dziewięć pomników i co najmniej jedna fontanna (zlokalizowana przed muzeum Ziemi Mińskiej).

W klasie obiektów *Kompleksy użytkowania terenu* (KUA_A) zabrakło wydzielenie obiektów należących do następujących klas: *osiedle mieszkaniowe*, *zespół szkół*, *zespół szpitalny czy zakład produkcyjny*, definiowanych w *Wytycznych technicznych TBD* (2008) i występujących na analizowanym terenie (rys. 3.4).

Ocena spójności logicznej

W odniesieniu do zbiorów danych TBD spójność logiczna jest rozumiana jako zgodności z logicznymi regułami struktury danych, atrybutów i związków, i oznacza brak wewnętrznych sprzeczności w zbiorze danych. Zgodnie z SKDBT w ramach oceny spójności logicznej należy sprawdzić warunek współliniowości i ciągłości obiektów, istnienie reguł przestrzennych wynikających z koncepcji TBD i opisanych w *Wytycznych Technicznych TBD* oraz poprawność segmentacji obiektów liniowych.

Badanie współliniowości zostało przeprowadzone dla obiektów liniowych klas *Odcinki jezdni* i *Budowle mostowe* (SKJZ_L – BBMO_L, rodzaj ‘m-most’) oraz *Odcinki jezdni* i *Budowle ziemne* – rodzaj nasyp (BBZM_L – SKJZ_L rodzaj ‘nasyp’). Do kontroli wykorzystano narzędzia selekcji programu ArcGIS, umożliwiające znalezienie obiektów, pomiędzy którymi zachodzą określone relacje przestrzenne, w tym przypadku są to relacje styczności.

Ciągłość obiektów powierzchniowych przeanalizowano dla obiektów klas *Tereny leśne lub zadrzewione* (PKLA_A) i *Obszary wód* (PKWO_A). Przeprowadzony test miał na celu wyodrębnienie obiektów nakładających się i został wykonany z wykorzystaniem operatorów i funkcji analiz przestrzennych.

Do oceny reguł przestrzennych zdefiniowanych w *Wytycznych Technicznych TBD* wybrano punkty adresowe (ARAD_P), które powinny być zlokalizowane w obrębie budynku (BBDA_A). Do weryfikacji warunku, tak jak poprzednio, wybrano operatory selekcji przestrzennej oraz operatory selekcji wg atrybutów.

Wyniki oceny spójności logicznej

Warunek współliniowości spełniły wszystkie mosty (7 obiektów klasy BBMO01 – *Budowle mostowe*, rodzaj most) i nasypy (3 obiekty klasy BBZM_L – *Budowle mostowe*, rodzaj nasyp). Na analizowanym obszarze zarówno mosty, jak i nasypy są usytuowane dokładnie w osi drogi (jezdni).

Nie zaobserwowano również przestrzennego nakładania się obiektów klas *Tereny leśne lub zadrzewione* (PKLA_A) i *Obszary wód* (PKWO_A). Wszystkie obiekty zgromadzone w tych klasach (7 obiektów) zachowują warunek ciągłości zdefiniowany w SKBDT.

Poprawność umieszczania punktu adresowego wewnątrz budynku wykazała, że co prawda wszystkie punkty adresowe (4601 obiektów) zlokalizowane są w obrębie budynku, niemniej jednak część z nich (15 obiektów stanowiących 0,33%) została przypisana do niewłaściwego budynku, np. budynku gospodarczego.

Ocena dokładności położenia obiektów

Na dokładność geometryczną obiektów TBD wpływają materiały źródłowe, poprawność interpretacji oraz dokładność wektoryzacji. Zgodnie z wytycznymi TBD błędy położenia obiektów nie powinny przekraczać 1,5 m. Dane pozyskane ze źródeł o większej dokładności niż w TBD powinny zostać poddane procesowi generalizacji. Odnosi się to przede wszystkim do map wielkoskalowych, z których pozyskiwane są np. dane o budynkach i granicach administracyjnych. Ogólna dokładność położenia obiektów w TBD wynosi 5 m dla pierwszej grupy dokładnościowej (całkowity średni błąd położenia obiektów) i 10 m dla pozostałych. Dokładność położenia, nazwana w TBD dokładnością geometryczną, należy do atrybutów obligatoryjnych wszystkich obiektów przestrzennych. Sprawdzone również wymagania dokładnościowe dotyczące wielkości budynków, które powinny mieć powierzchnię większą niż 40 m² i minimalną długość załamania/występu ściany budynku nie mniejszą niż 4 m. Przeprowadzona kontrola dokładności położenia obiektów obejmowała obiekty klasy *Budynki* (BBBD_A). Jako materiał referencyjny, w stosunku do którego przeprowadzono kontrolę, wykorzystano dane zgromadzone w egib.

Dodatkowo sprawdzono warunek minimalnej powierzchni i długości obiektów wprowadzanych do bazy TBD. Zgodnie w wytycznymi technicznymi (*Wytyczne TBD Cz.2 Załącznik 1 Zasady pozyskiwania*) dokładność przedstawiania opisu geometrycznego obiektów odpowiada mapie topograficznej w skali 1:10 000, a zatem minimalna odległość pomiędzy dwoma punktami pośrednimi nie może być mniejsza niż 2 m, a minimalna długość obiektu liniowego powinna wynosić 5 m. W odniesieniu do kompleksów pokrycia terenu wydzielane powierzchnie podlegają następującym kryteriom:

- minimalna szerokość wydzielanej powierzchni powinna być większa niż 10 m (odstępstwem mogą być obszary wód i terenów komunikacyjnych),
- minimalna powierzchnia powinna wynosić minimum 500 m² (wyjątek wody),
- dla obszarów wód wynoszą odpowiednio 5 m dla szerokości i 80 m² dla powierzchni.

Warunki minimalnej długości i powierzchni sprawdzono dla terenów leśnych (obiekty klasy PKLA_A) i obszarów wód (obiekty klasy PKWO_A), korzystając ze standardowych funkcji programu ArcGIS.

Wyniki oceny dokładności obiektów

Wyniki analiz nie pozwalają jednoznacznie stwierdzić czy lokalizacja wszystkich budynków (BBBD_A) została wykonana z wymaganą dokładnością. Ze szczegółowej analizy danych dotyczących 116 budynków, wybranych ze zbioru liczącego 7476 obiektów, wynika że przesunięcia w usytuowaniu budynku TBD względem budynku egib wahają się od 0,2 do 8,3 m. Średnia wartość przesunięcia budynku TBD względem jego odpowiednika w warstwie z egib wyniosła 1,6 m. Wartość ta mieści się w granicach błędu średniego dla danych zawartych w TBD. Jednak ocena przemieszczeń jest utrudniona ze względu na różne skale i poziom dokładności obydwu zbiorów.

Analiza wielkości powierzchni wykazała, że 363 budynki mają powierzchnię mniejszą niż 40 m², jednak 10 spośród nich znajduje się na granicy opracowania wobec czego ich geometria może być zniekształcona, dlatego w końcowym zestawieniu błędy obiektów będących na styku opracowania zostały pominięte. Za błędne uznano 353 budynki, co stanowi 4,72% i jest wartością prawie pięciokrotnie przekraczającą dopuszczalny procent błędnych obiektów (wg *Wytocznych Technicznych TBD* – 1%).

Sprawdzenie warunku minimalnej powierzchni i szerokości dla obiektów PKLA *Tereny leśne i zadrzewione* wykazało, że jedynie 4,04% (4 obiekty na 99) nie spełnia warunku minimalnej powierzchni. Natomiast kryterium długości spełniają wszystkie obiekty. Z kolei dla obiektów PKWO_A – *Obszary wód* oba warunki minimalnej długości i powierzchni zostały spełnione. Wśród 21 obiektów żaden nie miał mniejszej niż 8 m długości i mniejszej niż 80 m² powierzchni.

Ocena dokładności czasowej

W TBD informacje o aktualności i dacie utworzenia muszą charakteryzować obiekty wszystkich klas (tab. 3.2). Warunek ten sprawdzono wykorzystując operatory selekcji wg atrybutów.

Tabela 3.2. Wymagalność odniesień czasowych w TBD

Nazwa atrybutu	Typ danych	Wymagane ("not null")	Opis atrybutu, przykładowe wartości
X_AKTUALNOSC_G	DT	TAK	Stan aktualności geometrii obiektu
X_AKTUALNOSC_A	DT	TAK	Stan aktualności atrybutów obiektu
X_DATA_UTWORZENIA	DT	TAK	Data utworzenia obiektu
X_DATA_MODYFIKACJI	DT	TAK	Data modyfikacji obiektu

Wyniki oceny dokładności czasowej

Zbiory danych TBD dla Mińska Mazowieckiego zachowują spójność czasową, w zbiorach nie zaobserwowano błędnie wprowadzonych dat. Dаты utworzenia obiektu i dаты modyfikacji obiektu są takie same dla wszystkich klas.

Ocena dokładności tematycznej

Dokładność tematyczna zbioru danych przestrzennych Bazy Danych Topograficznych powinna być oceniana pod względem poprawności klasyfikacji, poprawności atrybutów nieilościowych i dokładności atrybutów ilościowych. Częściowo kontrole te są wykonywane standardowo i zamieszczane w *Raporcie pełnym*. *Raport* ten jest generowany podczas automatycznych kontroli SKBDT, które przede wszystkim dotyczą kontroli atrybutów obiektów i wykazów danych. *Raport* umożliwia ogólne i całościowe spojrzenie na zbiór danych i wychwycenie błędnych wartości we wszystkich polach tabel.

Wyniki oceny dokładności tematycznej

Raport pełny z kontroli zbiorów TBD dla badanego obszaru, wykazał wiele niezgodności. Dotyczyły one m.in. braku unikalności wartości atrybutu ID w tabeli lub występowania zależności pomiędzy atrybutami. Szczegółowa analiza danych wykazała, że błędy dotyczą obiektów występujących na granicy obszaru (dwa fragmenty lasu), które *de facto* są jednym obiektem, sztucznie podzielonym przez granicę opracowania (rys. 3.6).

3.5.3. Podsumowanie oceny jakości danych i dyskusja

W ocenie jakości danych przestrzennych wykorzystano dane otrzymane z państwowego zasobu danych geodezyjnych i kartograficznych, w tym zbiory danych Bazy Danych Topograficznych dla miasta Mińsk Mazowiecki, raporty z Systemu Kontroli TBD, dane z bazy ewidencji gruntów i budynków oraz mapy udostępniane w ramach geoportal.gov.pl. Sprawdzeniu podlegały błędy wykryte w kontroli automatycznej, zapisane w raportach z kontroli przeprowadzonej w ramach SKBDT oraz wybrane klasy obiektów. Kryteria, według których oceniano zasób zostały wybrane zgodnie z PN-EN ISO 19113 z 2009 roku – *Informacja geograficzna – Podstawy opisu jakości*. Do każdego z 5 kryteriów – kompletności, spójności logicznej, dokładności położenia, dokładności czasowej oraz dokładności tematycznej – zostały dobrane odpowiednie rodzaje analiz danych mające na celu ocenę jakości danych TBD. Ocena jakości umożliwiła określenie wiarygodności Bazy Danych Topograficznych dla wybranego obszaru, czyli stopnia zgodności danych z rzeczywistą sytuacją w terenie.

Wykaz przeprowadzonych kontroli oraz uzyskane wyniki zestawiono w tabelach 3.3 i 3.4.

1. Zasób danych przestrzennych dla Mińska Mazowieckiego wymaga uzupełnienia. W klasach obiektów poziomu drugiego OIPR_P – Obiekty przyrodnicze, BBHY – Budowle hydrotechniczne, BBIU_A – Inne urządzenia techniczne, BBUD_L – Umocnienia drogowe lub kolejowe, BBUW_L – Umocnienia wodne, pomimo wyróżnienia w bazie, nie zgromadzono żadnych obiektów w tabelach.
2. Dane przestrzenne w bazie są niekompletne i nie zawierają ważnych punktów orientacyjnych w terenie, takich jak kapliczki, krzyże czy fontanna.
3. Analiza kompletności klasy *Budynki* (BBBD_A) wykazała, że w TBD brak jest 142 obiektów, co stanowi 1,90% wszystkich budynków, ponadto zbiór zawiera 57 budynków nie zewidencjonowanych w egib. Dokładność położenia budynków określono na 1,6 m (średnie przesunięcie obiektów warstwy BBBD_A), co mieści się w granicach dopuszczalnego błędu. Ponad 4,7% budynków (353 budynki) nie spełnia kryterium wielkości powierzchni, a do 15 budynków gospodarczych (0,33% budynków) niepotrzebnie przypisano punkt adresowy (zamiast do budynku mieszkalnego).
4. Obiekty liniowe i powierzchniowe podlegające kontroli spełniły warunki spójności przestrzennej danych, określone w wytycznych technicznych.
5. Kompleksy użytkowania terenu nie zachowują ciągłości.
6. Warunku minimalnej powierzchni (poza budynkami) nie spełniły obiekty klasy tereny leśne (PKLA_A) – 4 obiekty (4,04% obiektów) oraz klasy kompleksy użytkowania terenu (KUA_A) – 2 obiekty (25%). Natomiast warunku minimalnej długości nie spełniły obiekty klasy odcinki jezdni o nawierzchni twardej (SKJZ_L) – 30 obiektów (1,78%).
7. Kontrole atrybutowe (m.in. zgodności kodów), których wyniki zostały opisane w raportach nie wykazały znaczących błędów. Dokładność tematyczna atrybutów obiektów zbioru została zachowana, co potwierdzają zestawy kontrolne przedstawione w raporcie.

Tabela 3.3. Wyniki oceny jakości TBD (arkusze N-34-140-A-c-3, N-34-140-A-c-4)

Elementy jakości	Rodzaj przeprowadzonej analizy	Wynik zgodności
Kompletność	skład zbioru danych	pozytywny
	wymagalność atrybutów	pozytywny
	unikalność atrybutów	pozytywny
	nadmiar obiektów/duplikaty	pozytywny
	skład zbioru warstwy BBBB_A	negatywny
	brak obiektów w warstwach	negatywny
Spójność logiczna	współliniowość obiektu	pozytywny
	ciągłość kompleksów pokrycia	pozytywny
	reguły przestrzenne	niejednoznaczny
	ciągłość i segmentacja obiektów liniowych	pozytywny
Dokładność położenia	przesunięcia obiektów	niejednoznaczny
Dokładność czasowa	spójność czasowa danych TBD	pozytywny
Dokładność tematyczna	kontrole atrybutowe	pozytywny
	minimalne powierzchnie i długości	negatywny

Tabela 3.4. Wyniki oceny dodatkowej TBD (arkusze N-34-140-A-c-3, N-34-140-A-c-4)

Element jakości	Rodzaj przeprowadzonej analizy	Oceniana warstwa	% obiektów spełniających kryteria
Kompletność	obiekty nie występujące w TBD	BBBD_A	98,10
	obiekty nie występujące w EGIB	BBBD_A	99,24
Spójność logiczna	współliniowość	BBMO_LSKJZ_L	100
		BBZM_LSKJZ_L	100
	ciągłość	PKLA_APKWO_A	100
	reguły przestrzenne	ARAD_PBBBD_A	99,67
	ciągłość i segmentacja	SKJZ_L	100
		SW	100
		SU	100
		SC	100
Dokładność położenia	pokrywanie się obiektów	BBBD_A i budynki Egib	87,94
Dokładność czasowa	spójność czasowa	wszystkie klasy obiektów TBD	100
Dokładność tematyczna	minimalna powierzchnia	PKLA_A	95,96
		PKWO_A	100
		KUAA_A	75
		BBBD_A	95,73
	minimalna długość	SKJZ_L	98,12

3.5. Znaczenie oceny jakości danych

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. Dlatego też, aby wykorzystywać systemy GIS w procesie decyzyjnym niezbędna jest nie tylko znajomość jakości danych, ale także modeli rozprzestrzeniania się błędów i umiejętność oceny wyników operacji przestrzennych.

Kluczowym zagadnieniem w ocenie jakości danych jest identyfikacja źródeł błędów. Większość błędów jest związana z przyjętym modelem świata rzeczywistego (semantycznym, logicznym i fizycznym), czyli z tym jakie obiekty i ich atrybuty zostaną zgromadzone oraz w jaki sposób zostaną zapisane w bazie danych. Ich wielkość zależy od oceny dokładności położenia obiektów i określenia atrybutów, a także precyzji zapisu danych. W celu uzmysłowienia sobie o jak dużych błędach mówimy, konieczne jest oszacowanie ich wielkości. Metody szacowania błędów różnią się w zależności od rodzaju błędów. Innymi metodami określa się błędy położenia, innymi błędy atrybutów ilościowych, a jeszcze innymi jakościowych. Błędy, jakimi obciążone są dane źródłowe, przenoszą się na dane przetworzone. Analiza przenoszenia się błędów z danych źródłowych do danych wynikowych jest określana mianem rozprzestrzeniania się błędów. Błędami różnego rodzaju obciążone są wszystkie dane geograficzne. Umiejętność oceny stopnia ryzyka przy podejmowaniu decyzji na podstawie danych obciążonych błędami nazywa się strategią zarządzania błędami.

W celu zminimalizowania liczby błędów w danych przestrzennych należy wykorzystywać tylko dane udokumentowane i pochodzące ze znanych źródeł. Przystępując do budowy bazy danych przestrzennych należy w każdym przypadku zdefiniować kryteria jakości, jakim mają odpowiadać dane. Należy pamiętać, że pozyskiwanie danych szczegółowych o wysokiej dokładności jest kosztowne i długotrwałe. Zmniejszenie szczegółowości i dokładności danych prowadzi do skrócenia czasu ich pozyskania i zmniejszenia kosztów, wpływa jednak na spadek dokładności analiz przestrzennych i wzrost niepewności w procesie podejmowania decyzji.

W celu zminimalizowania błędów danych przestrzennych i otrzymania jak najwyższej jakości produktów informatycznych należy:

- 1) kontrolować jakość danych na każdym etapie ich przetwarzania począwszy od wprowadzania danych,
- 2) unikać łączenia danych szczegółowych z danymi zgeneralizowanymi,
- 3) zrozumieć istotę danych i zasady ich przetwarzania,
- 4) analizować otrzymane wyniki pod kątem ich poprawności.

Literatura

- Bielecka E., 2009: Metadane i jakość danych przestrzennych. [W:] Modelowanie pojęciowe w projektowaniu i implementacji systemów geoinformacyjnych, Wydawnictwo IGIK, Warszawa, 97-106.
- Bielecka E., 2006: Systemy Informacji Geograficznej. Teoria i zastosowania. Wyd. PJJWSTK, Warszawa, 225 s.
- Draft Commission Regulation (EU) No .../..of [...] Implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services <http://inspire.jrc.ec.europa.eu/>, Brussels, 11.12.2009, D007474/02.

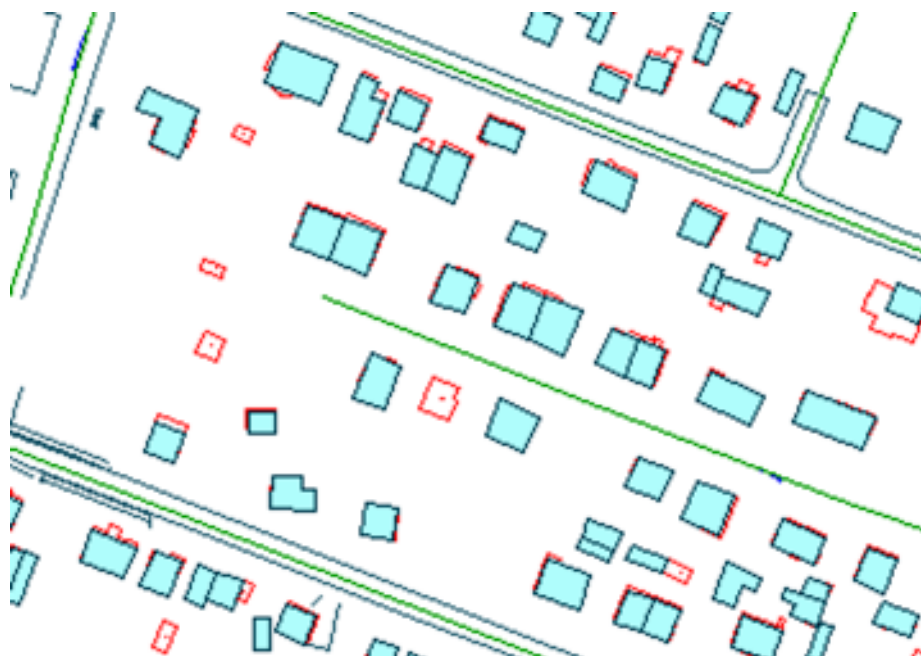
- Dyrektywa 2007/2/WE Parlamentu Europejskiego i Rady z dnia 14 marca 2007 r. ustanawiająca infrastrukturę informacji przestrzennej we Wspólnocie Europejskiej (INSPIRE) (L108/1).
- Gaździcki J., 2004: Internetowy Leksykon Geomatyczny. www.ptip.org.pl
- ISO, 2003a/Cor1:2005: ISO 19114:2005/AC:2006: Geographic information – Quality elevation procedures.
- ISO, 2003b/Cor1:2006: ISO 19115:2005/AC:2008: Geographic information – Metadata.
- ISO, 2002: ISO 19113:2005: Geographic information – Quality principles.
- ISO, 2006: ISO/TS 19138 Geographic information – Data quality measures.
- Kmieciak A., 2004: Analiza dokumentów ISO serii 19100 w zakresie metadanych i jakości danych geograficznych. Biuletyn Informacyjny Głównego Geodety Kraju, numer specjalny maj 2004: 23-36.
- Koncepcja Systemu Kontroli Bazy Danych Topograficznych TBD, Intergraph, 2006, dostępne http://www.bgwm.pl/pliki/tbd/SKBDT_Koncepcja_v1.27b.pdf
- Łabaj A., 2009: Kontrola jakości bazy danych topograficznych, materiały konf. SKP, Duszniki Zdrój 2009, dostępne <http://www.gislab.ar.wroc.pl/IIIZKSKP/Referaty/PALabajDuszniki2009.pdf>
- PN-EN ISO 19113:2009 Informacja geograficzna – Podstawy opisu jakości.
- Pachelski W., 2009: Modelowanie informacji geograficznej według znormalizowanej metodologii, [W:] Bielecka E. (red.), Modelowanie pojęciowe w projektowaniu i implementacji systemów geoinformacyjnych. Wydawnictwo IGIK, Warszawa, 7-22.
- Rozporządzenie Komisji Wspólnoty Europejskiej Nr 1205/2008 z dnia 3 grudnia 2008 r. w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie metadanych, (L 326/16).
- Wąsowska M., 2010: Ocena wiarygodności danych przestrzennych na przykładzie TBD. Praca magisterska na Wydziale Inżynierii Lądowej i Geodezji WAT, napisana pod kierunkiem dr hab. inż. E. Bieleckiej, Warszawa.
- Wytyczne techniczne Baza Danych Topograficznych – wersja 1.0 uzupełniona, Główny Urząd Geodezji i Kartografii, Warszawa 2008.

```

- <MechanizmKontroliPrzestrzennych id="topo_e3" name="Kontrole konstrukcji geometrii"
  coordSystemFilePath="{APP_PATH}\csf\puwg1992.csf" dataLotElementDefinitionId="pd_topo">
- <Konektory>
  <GeoMediaDataConnector connectionName="TOPO" location="{CONV_PATH}\topo.mdb"
    type="AccessRO.GDatabase" />
  <GeoMediaDataConnector connectionName="RANGE" location="{CONV_PATH}\range.mdb"
    type="AccessRO.GDatabase" />
</Konektory>

```

Rys. 3.2. Fragment pliku xml szablonu kontroli TBD (źródło: Labaj, 2009)



Rys. 3.3. Sprawdzenie kompletności budynków zgromadzonych w TBD; czerwonym obrysem oznaczono budynki zgromadzone w EGIB, natomiast wielobokami wypełnionymi kolorem zielonym – budynki w TBD

ID	RODZAJ	NAZWA	SHAPE_Length	SHAPE_Area
150004	oczyszczalnia ścieków	<Null>	390,003082	4040,25325
150005	cmentarz	<Null>	460,257514	11128,63255
150008	cmentarz	<Null>	1681,845953	91452,13325
150007	wysypisko odpadów	<Null>	524,861099	14716,3614
150009	park	Park Dermalowiczów	2350,280291	241521,894
150035	park	<Null>	597,718935	10059,23115
150036	elektrociepłownia	<Null>	269,131331	4086,9411
150038	oczyszczalnia ścieków	<Null>	621,464111	19022,5227

Rys. 3.4. Obiekty warstwy *Kompleksów użytkowania terenu*, TBD dla Mińska Mazowieckiego

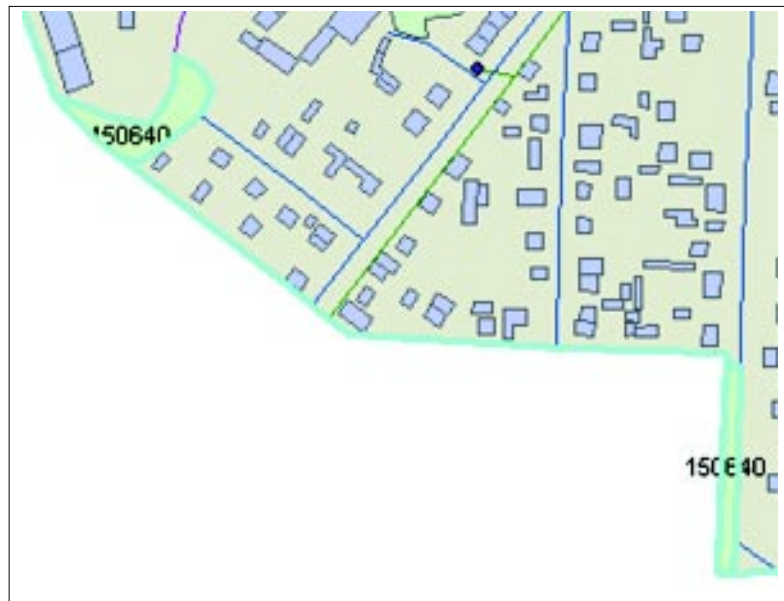


a



b

Rys. 3.5. Mińsk Mazowiecki:
a – pomnik bohaterów II wojny światowej,
b – pomnik 1Pułku Lotnictwa Myśliwskiego
(źródło: http://pl.wikipedia.org/wiki/Mi%C5%84sk_Mazowiecki)



Rys. 3.6. Weryfikacja unikalności wartości atrybutu ID dla PKLA

Wprowadzenie

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

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

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

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

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

Marek Baranowski
Elżbieta Bielecka

Rozdział 1. Modelowanie informacji geograficznej: Podstawy

Chapter 1. Geoinformation modeling: Fundamentals

prof. dr hab. inż. Wojciech Pachelski

Wydział Inżynierii Lądowej i Geodezji, Wojskowa Akademia Techniczna

WojciechPachelski@aster.pl

Rozdział 2. Transformacja polskich danych przestrzennych do modeli INSPIRE

Chapter 2. Transformation of Polish spatial data to INSPIRE models

dr hab. Janusz Michalak

Wydział Geologii, Uniwersytet Warszawski

J.Michalak@uw.edu.pl

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

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

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

dr hab. inż. Elżbieta Bielecka, prof. WAT

Instytut Geodezji i Kartografii

Wydział Inżynierii Lądowej i Geodezji, Wojskowa Akademia Techniczna

elzbieta.bielecka@igik.edu.pl

MODELOWANIE DANYCH PRZESTRZENNYCH

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

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

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

Streszczenie

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

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

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

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

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

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

SPATIAL DATA MODELING

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

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

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

Abstract

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

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

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

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

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

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

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

Spis treści

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