



Mgr inż. Karol KAIM

Prezes Zarządu – Dyrektor Generalny Okręgowego Przedsiębiorstwa Geodezyjno-Kartograficznego OPGK Sp. z o. o. w Gdańsku. Absolwent Wydziału Geodezji i Kartografii Politechniki Warszawskiej, geodeta uprawniony w zakresie 1 i 2, klasyfikator gruntów. Wykonawca licznych projektów modernizacji EGİB, założenia BDOT i GESUT, weryfikacji i konwersji danych. Członek SGP Oddział w Gdańsku.

## GML – Geodane Można Lubić, czyli dobry czy zły jest plik GML

Od wielu lat toczy się w środowisku geodezyjnym dyskusja na temat wymiany danych opracowywanych przez geodetów oraz roli informatyki w rozwoju geodezji.

Z przykrością stwierdzam, że coraz częściej dają się słyszeć głosy, jakoby wymianą danych geodezyjnych, bazami danych typu EGİB, BDOT czy GESUT powinni się zajmować specjaliści innych profesji niż geodezja, np. informatyki lub geoinformatyki. No cóż, nasz zawód, jak wszystko dokoła, zmienia się dynamicznie, ale czy to oznacza, że powinniśmy oddać to pole kolegom od komputerów? Moim zdaniem stanowczo nie.

Aby nie przegrać walkowerem, postanowiliśmy w OPGK w Gdańsku zaznajomić się bliżej z plikami GML i oddać, a właściwie zwrócić, bazy danych geodetom.

Jako, że jesteśmy praktykami, nasze podejście miało na celu produkcyjne wykorzystanie. Zależało nam, aby tematem GML zajęli się geodeci – osoby mające wieloletnie doświadczenie, choć niepotrafiący programować ani pisać skryptów w SQL.

Początki były trudne. Okazało się szybko, że raporty z walidatorów, dostępnych na rynku, są praktycznie nieczytelne dla praktyków geodetów, a interpretacja danych sprawia kłopot programistom. Stąd, podobnie jak w całej Polsce, tak i w naszym przedsiębiorstwie powstało przekonanie, że GML jest „trudny”, pliki „złe” i tak naprawdę to nie bardzo wiadomo, dlaczego nie chce nam to działać. Programiści twierdzili, że dane są złe i nie da się ich pokazać w GML, a geodeci, że to GML jest zły i informatycy powinni bardziej się postarać, by się dostosować do specyfiki naszego zawodu.

Rozwiązaniem okazał się „reset” całego projektu i powrót do tradycyjnego podejścia. Z merytorycznego punktu widzenia format GML jest jedynie formatem wymiany danych. Tak więc można go przyrównać do opakowania na czas przeniesienia danych. Przy takim podejściu potrzebujemy informatyków, aby pomogli nam zapakować dane lub, jeśli mamy gotowy plik GML, rozpakować je. Ale to my (geodeci) wiemy, jaka jest wartość danych, na co trzeba zwrócić szczególną uwagę oraz co trzeba zapakować, by przenoszenie nie uszkodziło cennego ładunku.

W wyniku naszego projektu powstało oprogramowanie o nazwie „GML Explorer”, którego zadaniem jest rozpakowywanie danych zapisanych w formacie GML tak, aby w czytelnej dla użytkownika postaci można było przeglądać i analizować. W ten sposób, dzięki pracy informatyków, otworzyła się możliwość analizowania danych przez geodetów – specjalistów, którzy wiedzą, co analizują i potrafią zapisać dane interpretować.

Okazało się szybko, że „nie taki diabeł straszny (...)” i że rozwój informatyki nie musi oznaczać dla naszego zawodu zagłady, ale wręcz przeciwnie, jeśli podejmiemy do tego z otwartością, to wykorzystanie nowych technologii informatycznych może nasz zawód wzbogacić.

Osoby, zajmujące się tym tematem naukowo, moje przemyślenia zapewne uznają za zbyt trywialne. Liczę jednak na wyrozumiałość, z uwagi na to, iż jestem w tej dziedzinie praktykiem i skupiam się na użytkowym wykorzystaniu przedstawianych treści.

Starając się zrozumieć czym jest UML, XML, GML, XSD i inne tego typu „nieładzkie” skróty, które wkroczyły agresywnie do geodezji, przeglądałem mozolnie dostępne opracowania oraz starałem się zasięgnąć wiedzy od osób zajmujących się informatyką, uznając, że do tej branży należą te pojęcia.

Teraz uważam, że myliłem się traktując te zagadnienia jako informatyczne. Od jakiegoś czasu staram się przekonywać, że jest to geodezja i to my (geodeci) jesteśmy najbardziej kompetentni, aby tymi zagadnieniami się zajmować.

Całe zagadnienie wymiany informacji, modeli, plików GML itd., proponuję dla lepszego zrozumienia przyrównać do języka (np. polskiego), który służy do komunikowania się, ale, aby komunikujący się rozumieli, treść komunikatu rządzi się pewnymi prawami.

Na początek kilka pojęć – zaczynając zgodnie z geodezyjną zasadą od ogółu do szczegółu:

*UML – Unified Modeling Language, czyli Ujednolicony Język Modelowania – to formalny język pozwalający na budowanie modeli mających na celu odzwierciedlić pewien fragment rzeczywistości. Można przyrównać UML do rodziny języków, które w podobny sposób opisują rzeczywistość, mają podobną gramatykę i składnię (np. języki indoeuropejskie).*

*XML – Extensible Markup Language, Rozszerzalny Język Znaczników, to uniwersalny język formalny przeznaczony do reprezentowania różnych danych w strukturalizowany sposób.*

Można XML przyrównać do zestawu wykorzystywanych słów oraz reguł, jak budować zdania.

*GML – Geography Markup Language, Język Znaczników Geograficznych, jest odmianą języka XML do opisu danych przestrzennych.*

Tak więc GML można przyrównać do konkretnego języka z grupy (np. język polski) – który ma swoją zdefiniowaną gramatykę, interpunkcję, słownictwo i inne zasady, których należy przestrzegać, aby poprawnie się w nim komunikować.

Do przekazywania danych XML fizycznie potrzeba dwóch plików:

- z rozszerzeniem **XSD** – definiuje strukturę danych przekazywanych w formie obiektów.

Można go przyrównać do słowniczka oraz reguł gramatyki objaśniających jak będą zapisane konkretne pojęcia oraz jak je rozumieć. W naszym przypadku jest to wzorzec znajdujący się w przepisach, a więc nie ma konieczności przekazywania go każdorazowo z danymi.

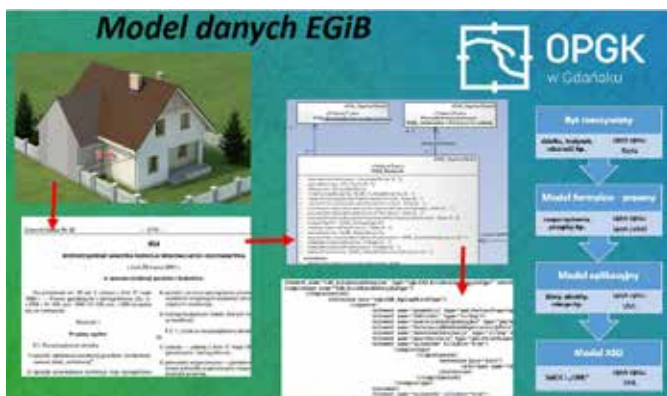
- z rozszerzeniem **XML (GML)** – zawiera konkretne wartości przekazywanych danych.

Czyli: konkretny tekst zapisany w konkretnym języku (np. polskim) (XML\GML) zgodnie z jego zasadami (gramatyką) pochodzącym z grupy języków (UML), zawierający merytoryczny opis konkretnych obiektów rzeczywistych przy pomocy zdefiniowanych słów i pojęć (XSD).

Geodeci od zawsze zajmowali się opisem rzeczywistości za pomocą aktualnie dostępnych środków technicznych. Wykorzystywaliśmy papierowe mapy, rejestry, kartoteki, opisy, potem mapy na folii, pliki komputerowe, a teraz bazy danych. Niezależnie od nośników i zaawansowania technicznego, staraliśmy się odzwierciedlić byty rzeczywiste (obiekty przestrzeni) na podstawie modelu formalno-prawnego (przepisy) w postaci możliwej do reprezentacji, analiz itd.

Tak więc wprowadzenie GML jako narzędzia w naszym zawodzie potraktować można jak nową instrukcję K1 – która określa, w jaki sposób przedstawić obiekty rzeczywiste na mapie lub w bazie danych. Oczywiście GML to bardziej zaawansowane rozwiązanie. Daje większe możliwości, ale wiąże się z nim szereg dodatkowych wyzwań takich jak: model jakości da-

nych (MJD), walidacji, implementacja do baz powiatowych, wymiana danych, przekazywanie danych inicjalnych oraz różnicowych itd.



Zagadnienia te są niezwykle ważne, ale rozważając je powinniśmy zawsze pamiętać, aby nie stracić z oczu danych, czyli treści jakie chcemy przenosić tak, by opakowanie nie stało się ważniejsze od zawartości.

Z naszej praktyki jako weryfikatora danych w projekcie ZSIN faza I wynika, że często dyskusja sprowadza się do wymiany argumentów, czy GML jest „dobry” czy „zły” oraz czy da się jakieś dane przenieść przy pomocy tego formatu czy nie. Tu niestety odpowiedź jest bardziej skomplikowana, ponieważ (korzystając z poprzednich porównań do języka) można zastanawiać się, czy jakaś wypowiedź w języku polskim (np. opis przedmiotu) jest dobra czy zła. W GML jest podobnie – zdarzają się różne błędy, czasami ortograficzne czy gramatyczne, niektóre z nich to tylko pomyłki, ale niektóre wręcz uniemożliwiają odczytanie lub zrozumienie tekstu. Czasami opis jest poprawny językowo, ale ma błędy merytoryczne. Podobnie z plikami GML, mogą być poprawne pod względem syntaktycznym, ale nieść błędne informacje. Tak więc, aby odpowiedzieć na pytanie, czy plik GML jest dobry, trzeba określić Model Jakości Danych. Jest to zestaw warunków, jakie dane muszą spełniać, aby zostać uznane za poprawne. W modelu jakości danych określić należy, jakie błędy traktowane będą jako krytyczne (których występowanie stanowić będzie powód negatywnej oceny pliku), a jakie będą dopuszczalne lub traktowane jedynie jako ostrzeżenia. Przecież inny model jakości danych zastosujemy do danych po zakończeniu procesu modernizacji ewidencji gruntów i budynków, a inny, jeżeli przedmiotem opracowania była np. weryfikacja jedynie oznaczeń użytków gruntowych.

Nie w budowaniu formatu lub czytaniu pliku GML, ale właśnie w definowaniu modelu jakości danych oraz interpretacji uzyskanych wyników walidacji danych jest miejsce dla geodetów. To my mamy wiedzę i doświadczenie, aby określić czy np. brak adresu budynku, nr KW albo daty zakończenia dzierżawy jest błędem, czy wynika z uwarunkowań powstania danych. Tak więc potrzebujemy narzędzi informatycznych, ale są to tylko narzędzia, a cała odpowiedzialność za wynik analizy spoczywa na obsługującym narzędzie geodecie.

Aby umożliwić tę analizę, narzędzie musi spełniać kilka warunków. Przede wszystkim raporty muszą być czytelne i odnosić się do przepisów prawa i konkretnych warunków modelu jakości danych. Narzędzie powinno też wizualizować dane oraz błędy w postaci mapy.



Podjęcie takie pozwoliło nam przestać zajmować się formą i skupić na treści danych.

## GML w projekcie ZSIN

Nasze doświadczenia z GML wynikały z faktu uczestnictwa w weryfikacji danych ZSIN faza I. W tej chwili w realizacji jest faza II projektu, ale nadal żywa jest dyskusja nt. zasilenia bazy centralnego repozytorium danymi z baz powiatowych. Mimo, że Główny Urząd Geodezji i Kartografii udostępnił narzędzie do walidacji plików GML, to kwestia poprawności danych oraz możliwości ich migracji do centralnego repozytorium ZSIN budzi nadal wiele pytań i kontrowersji.

Po pierwsze należy przypomnieć, że walidując pliki GML oceniamy je pod kątem:

- 1) występowania błędów niezgodności ze schematem XSD (które stosując poprzednie porównanie można potraktować jak błędy gramatyczne). Występowanie takich błędów dyskwalifikuje plik. Zgodność ze schematem XSD wynika wprost z przepisów prawa.
- 2) występowania błędów wynikających z założonego modelu jakości danych (można je porównać do poprawnie napisanego tekstu, który zawiera niepoprawną lub niepełną informację).

W ramach projektu ZSIN zamawiający stworzył MJD i to wypełnienie warunków tego modelu jest warunkiem przyjęcia danych do centralnego repozytorium.

W udostępnionej przez GUGiK aplikacji do walidacji plików dla ZSIN wykazywane błędy podzielono na kategorie: krytyczne, niekrytyczne, ostrzeżenia i informacje. Błędy krytyczne to takie, które uniemożliwiają zacytowanie danych do centralnego repozytorium. Do najczęściej występujących błędów krytycznych należą:

- braki atrybutów obiektów (statusy osób, płeć osoby, funkcje budynków itp.),
- błędne atrybuty, zły format danych, niewłaściwy skład OFU, OZU OZK itp.,
- niepełne dane (braki adresów dla lokali lub powielone adresy),
- działki i budynki tylko na mapie (brak części opisowej obiektów),
- błędy geometrii obiektów,
- braki adresów (mi.in. lokali),
- braki powiązań jednostek rejestru budynków z jednostkami rejestru gruntów,
- puste powierzchnie pomieszczeń przynależnych do lokali,
- brak lub błędne grupy rejestrowe,
- brak dat, od których obowiązuje umowa dzierżawy,
- błędy powiązań między obiektami (budynek w wielu obrębach, powiązania między obiektami różnych gmin),
- niekreślona przestrzeń nazw,
- różnica w datach utworzenia zmiany i daty modyfikacji obiektu, który podlegał tej zmianie.

Kłopot z wyeliminowaniem błędów krytycznych wynika, moim zdaniem, po pierwsze z trudnym do zrozumienia dla geodetów raportowaniem przez udostępniony walidator np.:

„urn: gugik: specyfikacje: gmlas: ewidencjaGruntowBudynkow: 1.0: licznikUdzialuWNieruchomosciWspolnej” jest nieprawidłowy. Wartość „jest nieprawidłowa przy uwzględnieniu jego typu danych „nonNegativeInteger” – Ciąg” nie jest prawidłową wartością Integer.

**Szczegóły weryfikacji:**

Export.gml: linia 7691185, pozycja 2: Element 'urn:gugik:specyfikacje:gmlas:ewidencjaGruntowBudynkow:1.0:rodzajLokalu' jest nieprawidłowy. Wartość "" jest nieprawidłowa przy uwzględnieniu jego typu danych 'urn:gugik:specyfikacje:gmlas:ewidencjaGruntowBudynkow:1.0:EGB\_RodzajLokaluType' - Błąd ograniczenia Enumeration.  
 Export.gml: linia 9233047, pozycja 2: Element 'urn:gugik:specyfikacje:gmlas:ewidencjaGruntowBudynkow:1.0:statusBudynku' jest nieprawidłowy. Wartość "" jest nieprawidłowa przy uwzględnieniu jego typu danych 'urn:gugik:specyfikacje:gmlas:ewidencjaGruntowBudynkow:1.0:EGB\_StatusBudynkuType' - Błąd ograniczenia Enumeration.

*Przykład raportu z walidatora GUGIK*

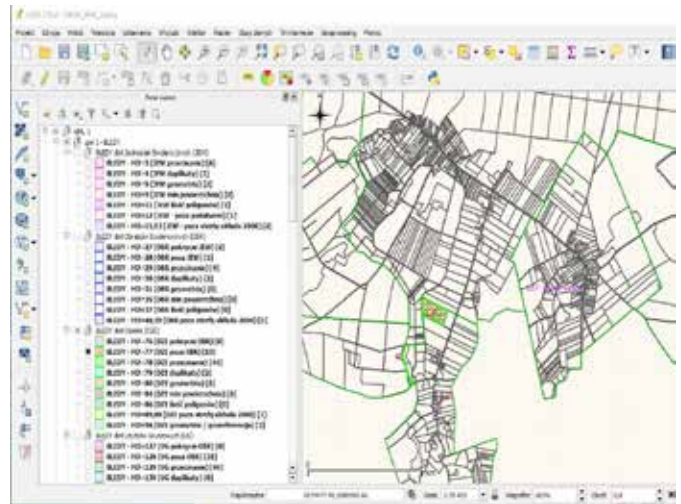
Błąd – braku powiązań pomiędzy udziałem z rejestru lokali a udziałem z rejestru gruntów (tzw. prawa związane).

*Element element „geometria” w przestrzeni nazw, [...] ewidencjaGruntowBudynkow: 1.0. ma nieprawidłowy element podrzędny element „MultiCurve*

Są to błędne rodzaje geometrii, najczęściej schodów dla budynków, które wg schematu powinny być zamkniętym poligonem, a w bazie są wielolinią. Inny przykład błędu (brak statusu budynku):

Drugim bardziej, moim zdaniem, ważnym problemem jest, które błędy wynikające z założonego modelu jakości danych uznano za błędy „krytyczne”. W przepisach nie ma definicji, który błąd jest krytyczny, a który to jedynie ostrzeżenie. O wadze błędu mówi przyjęty MJD, czyli warunki, jakie postawił wobec danych zamawiający. W tej sytuacji można się zastanawiać, czy faktycznie tak „krytycznym” z punktu widzenia merytorycznej wartości bazy danych EGIB błędem jest brak daty rozpoczęcia dzierżawy – zwłaszcza, że brak daty jej zakończenia błędem krytycznym już nie jest. Czy faktycznie błędem krytycznym powinien być brak adresu dla lokalu, grupy rejestrowej lub powierzchni pomieszczeń przynależnych do lokalu.

Analizując pliki GML w stworzonym przez OPGK w Gdańsku walidatorze, staraliśmy się zmienić wygląd raportów tak, by był on czytelniejszy dla użytkownika oraz pokazać dane i występujące błędy na mapie.



*Przykład raportów z walidatora OPGK w Gdańsku*

mają mniejsze znaczenie. Może warto byłoby skupić się nad sprawami ważniejszymi jak poprawność topologii obiektów (działki nachodzące wzajemnie na siebie nie są traktowane jako błąd krytyczny w ramach ZSIN), wyeliminowanie map analogowych i centroid, poprawa wiarygodności danych, niż uzupełnianie „na siłę” wszelkich atrybutów, ponieważ ich brak został potraktowany jako błąd krytyczny. Takie uzupełnienia bez wymaganych materiałów i źródeł, zamiast poprawić, mogą pomniejszyć wartość naszych baz przez obniżenie wiarygodności.

Zapewne wiele osób będzie miało odmienne zdanie, ale ważne jest byśmy w dyskusjach czy dobry, czy zły jest plik GML, skupili się jednak na wartości merytorycznej danych zapisanych w formacie GML, a nie na samym formacie, który jest tylko opakowaniem dla naszej cennej zawartości.

Zainteresowanych naszym podejściem do danych zapisanych w formacie GML – oprogramowaniem do ich czytania, walidacji oraz analizowania zapraszam na stronę [www.opgk.gda.pl/GML](http://www.opgk.gda.pl/GML), gdzie prezentujemy w sposób szczegółowy nasze rozwiązania.

Walidacja plików oraz dobranie właściwego modelu jakości danych nie jest jedynym wyzwaniem związanym z wprowadzeniem plików GML w geodezji. Nadal otwartą pozostaje kwestia definiowania modelu jakości danych dla konkretnego opracowania, przypisania konkretnym rodzajom błędów wagi zasilania baz danych (inicjalnego i różnicowego), przekazywanie w GML danych niestandardowych, wymiany plików GML z branżami oraz wiele innych. O problemach tych – w kolejnych wydaniach *Przeglądu Geodezyjnego*.

**EGB\_LokalSamodzielny**

Kontrola poprawności lokali

IdLokalu	SzczegółyBłędu	StartWersjaObiekt
321703_4.0001.230_BUD_2_LOK	atrybut liczbaPomieszczenPrzynaleznych powinien być równy liczby relacji obiektów EGB_FomieszczeniePrzynalezneDoLokalu do obiektu EGB_LokalSamodzielny	26.11.2008 11:32:40
321703_4.0001.296_BUD_2_LOK	liczbaIzb powinna być liczbą całkowitą dodatnią	15.07.2009 10:22:40
321703_4.0001.32_BUD_7_LOK	pow.UzytkowLokalu powinna być zapisana z dokładnością do 2 miejsc po przecinku	07.04.2010 12:35:21
321703_4.0001.1792_BUD_8_LOK	atrybut liczbaPomieszczenPrzynaleznych powinien być równy liczby relacji obiektów EGB_FomieszczeniePrzynalezneDoLokalu do obiektu EGB_LokalSamodzielny	24.03.2011 16:34:45
321703_4.0001.136_BUD_5_LOK	Dla kondygnacji nadziemnych liczba całkowita dodatnia od 1 (parter – pierwsza kondygnacja nadziemna) do n, dla kondygnacji podziemnych liczba całkowita ujemna od -1 (pierwsza kondygnacja podziemna) do -n	16.05.2012 11:25:00
321703_4.0001.33_BUD_1_LOK	liczbaIzb powinna być liczbą całkowitą dodatnią	18.03.2011 15:01:27
321703_4.0001.3_BUD_3_LOK	liczbaIzb powinna być liczbą całkowitą dodatnią	18.03.2011 13:55:18
321703_4.0001.1792_BUD_15_LOK	atrybut liczbaPomieszczenPrzynaleznych powinien być równy liczby relacji obiektów EGB_FomieszczeniePrzynalezneDoLokalu do obiektu EGB_LokalSamodzielny	24.03.2011 16:48:31

Ważniejsza jednak wydaje się nam dyskusja nad przyjętym dla danych modelem jakości. Zbyt pochopne przyjęcie za błąd krytyczny braków, dla których nie ma w starostwie materiałów, powoduje negatywną ocenę bazy – dane zapisane w pliku nie spełniają wtedy modelu jakości danych, więc cała partia danych nie może być zaakceptowana.

Uważam, że jako geodeci nie powinniśmy oddawać tego pola informatykom, bo to my wiemy, co jest ważne w bazach danych i jakie elementy