

Jak dojść z danymi do plików GML i pozostać przy zdrowych zmysłach

Transformacja z głową

Na początku muszę zaznaczyć, że są to tylko i wyłącznie moje przemyślenia, niepoparte przez nikogo i niepopierające nikogo, więc na trochę sobie pozwoliłem. A te kłopoty z pozostaniem przy zdrowych zmysłach przy dochodzeniu do plików GML (przynajmniej u niektórych) można z czystym sumieniem zwalić na Unię Europejską. To wszystko przez nich!

Zenon Parzyński

To oni uchwalili dyrektywę INSPIRE i wszystkie państwa członkowskie zostały zmuszone do jej implementacji! Bo czy u nas ktoś wpadłby na pomysł, aby:

- Ustanowić jedną metodologię tworzenia modeli baz danych oraz język (UML) jej zapisu?

- Wszystkie modele były zharmonizowane ze sobą (w stopniu jak najwyższym)?

- Wszystkie zbiory danych posiadały zdolność do interoperacyjności?

Może tak, a może nie. Trudno powiedzieć, jak mogłoby być, ale do tej pory w „naszej” geodezji było krucho nawet z ustanowieniem standardów, a co dopiero mówić o wprowadzeniu np. jednego standardu wymiany danych dla wszystkich ich rodzajów!

Ale to mamy już za sobą. Przed nami jest „tylko” implementacja dyrektywy INSPIRE, której kulminacją ma być wymiana danych między różnymi systemami za pomocą plików GML (*Geography Markup Language*) bez potrzeby ingerencji człowieka (czyli bez konieczności przekształcania danych do innej postaci – zostaje tylko wybrać, które dane będziemy wysyłać). A do tego prowadzi uciążliwa w wielu przypadkach droga.

Na początku zostały opracowane i wydane ustawy (np. *o infrastrukturze informacji przestrzennej*) i rozporządzenia opracowane w Głównym Urzędzie Geodezji i Kartografii. W rozporządzeniach m.in. zdefiniowano struktury baz danych w postaci modeli UML (*Unified Modeling Language*) oraz te same modele, ale zapisane w postaci schematów aplikacyjnych GML. To właśnie te schematy w GML (często jest używane dla nich oznaczenie XSD, *XML Schema Definition*) stanowią standardy wymiany danych.

Teraz w zasadzie pozostaje już tylko doprowadzić do sytuacji, by te pliki GML

z danymi „fruwaly” po sieci od nadawcy do adresata. By z jednej strony człowiek był potrzebny tylko do ich wysłania (czyli wyboru, które dane, i kliknięcia odpowiedniej opcji w aplikacji), a z drugiej – do odebrania, czyli kliknięcia, dzięki któremu pliki te zostaną wczytane do aplikacji, ale nie do bazy danych! Po wczytaniu wypadałoby te dane sprawdzić (nie tylko formalnie, ale też merytorycznie) i dopiero, gdy okażą się dobre, zapisać do bazy danych.

Kto będzie miał tę dużą przyjemność dojścia do plików GML? Wszyscy ci, którzy z mocy prawa mają w swoich obowiązkach gromadzenie danych przestrzennych o strukturze zdefiniowanej w jednym z opublikowanych rozporządzeń. Najwięcej kłopotów będzie z danymi na samym „dole”, czyli najbardziej szczegółowymi (BDOT500, EGIB, GESUT itp.), m.in. dlatego, że jest ich najwięcej. Dla części danych ten kłopot spadnie na GUGiK (dotyczy to np. fotogrametrii). Kto zaś personalnie za to odpowiada? Wiadomo – szef! Czyli starosta, prezydent miasta, wojewoda itp. Należy pamiętać, że zaradny wódz zwał winę na szefa od geodezji – czasem będzie miał rację

Kolejne zasadnicze pytanie brzmi: Czy chodzi tylko o pliki GML? Według mnie – nie tylko. Pliki GML powinny powstać z poprawnej bazy danych, a żeby mieć taką bazę, trzeba najpierw uporządkować posiadane obecnie dane, przepisać je do nowej struktury (lub ewentualnie nie przepisywać). Te wszystkie kroki są celami pośrednimi w całym procesie dochodzenia do celu ostatecznego, którym – jak już wyżej było powiedziane – jest interoperacyjność, czyli wymiana danych w plikach GML przez internet. Ma być tak, że po zgłoszeniu pracy przez internet geodeta tą samą drogą dostanie dane w postaci „n” plików GML i w takiej samej postaci (tylko uzupełnionej treścią swojej pracy) będzie zwracał dane do ośrodka. W pojedynczym pliku będą dane z jednej bazy, dlatego jeśli praca będzie obejmować dane

np. z czterech różnych baz, geodeta otrzyma dane w czterech plikach.

Droga nie jest straszliwie długa, ale na pewno ma parę ostrych zakrętów i składa się z kilku etapów.

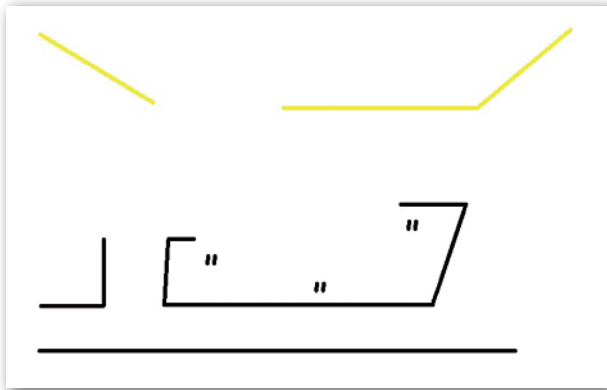
• 1. Struktura posiadanych danych

Konieczne należy poznać strukturę i format posiadanych danych: jakie obiekty znajdują się w bazie i jakimi atrybutami są opisywane, jakie relacje łączą te obiekty, w jakich formatach można „wydostać” dane z bazy. Spełnienie tego warunku jest niezbędne. Najczęściej dane znajdują się w relacyjnych bazach danych z tabelami. Ideałem jest jedna tabela odpowiadająca jednemu rodzajowi obiektu, ale bardzo często na obiekt składa się kilka tabel. Najczęściej pewne dane (głównie GESUT i BDOT500) mamy w postaci mapy, czyli w zasadzie nie ma bazy danych. Te informacje o strukturze i postaci danych będą podstawą opracowania koncepcji ich transformacji do postaci docelowej.

Transformację można zaplanować w różny sposób. Są programy, które pozwalają zdefiniować przejście z jednej struktury danych na inną, np. na podstawie dwóch modeli UML lub na podstawie danych w postaci XML (w październiku ub.r. w Paryżu odbyły się warsztaty poświęcone temu zagadnieniu). W każdym przypadku musi powstać tabela „tłumacząca”, która definiuje przejście od jednej struktury do drugiej. Do jej utworzenia niezbędny jest człowiek – maszyna sama tego nie robi.

Część obiektów w nowej strukturze będzie taka sama, dla części obiektów zmieni się opis atrybutami (niektórych w nowej strukturze nie będzie, pojawiają się za to nowe). W innych przypadkach trzeba będzie łączyć pewne elementy kilku „starych” obiektów, by utworzyć jeden „nowy”, lub z jednego „starego” zrobić kilka „nowych”. Takich różnych konfiguracji będzie sporo, ale każdą trzeba rozpatrzyć osobno i odpowiednio zdefiniować transformację w tabeli.

Wykorzystując do utworzenia tabeli „tłumaczącej” odpowiednie oprogramowanie, można potem transformację danych do nowej struktury przeprowadzić automatycznie. Jest tu jednak pewne „ale”.



Przykłady danych, które mogą zawierać błędy

• 2. Wartość posiadanych danych

„Wartość” należy w tym przypadku rozumieć jako stopień zgodności danych ze strukturą, w jakiej są zapisane, i z logiką (jak coś jest obszarem, to powinno być ograniczone krzywą/lamaną zamkniętą!). Obecnie często brakuje wartości niektórych atrybutów, granice obiektów nie są kompletne, brakuje danych o dokładności pewnych elementów itp. Na rysunku powyżej jest przedstawiony przewód gazowy w dwóch kawałkach oraz w dolnej części trawnik, a może raczej fragment granicy trawnika. Z kolei oba kawałki przewodu gazowego wydają się stanowić części jednej całości, ale równie dobrze mogą to być fragmenty dwóch przewodów. Dla niektórych rzeczą naturalną byłoby dołożyć (dorysować) te brakujące elementy, ale wtedy można postawić pytanie: **A po co w ogóle mierzyć? – przecież to jest kosztowne i kłopotliwe!** Jeśli odpowiedź na pytanie będzie następująca (dla mnie oczywista): nic nie dorysowujemy, trzeba najpierw sprawdzić w terenie, jak jest (bo może być już zupełnie inaczej) i domierzyć to, czego brakuje. Gdy są błędy/braki danych, mogą wystąpić poważne kłopoty przy próbach automatycznej transformacji.

Część danych nie jest zapisana w bazie danych, a tylko widoczna na mapie (analogowej, rastrowej czy wektorowej). Najczęściej taka sytuacja występuje dla danych GESUT i BDOT500. Wtedy bazy trzeba utworzyć i nie ma tu mowy o jakiejś transformacji danych. Szczególnie wtedy, gdy nie ma bazy danych, a mapa była analogowa, dane mogą być niekompletne (porównaj rysunek). W nowej strukturze takich obiektów nie można zapisać. Takich przypadków zapewne będzie wiele – im jest ich więcej, tym „wartość” danych jest mniejsza.

By w miarę rozsądnie zaplanować kolejne działania, konieczne trzeba być świadomym struktury, formatów oraz „wartości” danych. I stawiam tezę, że to uporządkowanie, „wyczyszcze-

nie” błędów powinno być jednym z głównych celów różnych działań podejmowanych teraz! Dokładnie tak, jak napisałem na początku.

• 3. Aplikacja

Kolejny krok to odpowiedź na pytanie: Jaka aplikacja? Ta obecna, tylko w nowszej wersji dostosowanej do nowych rozporządzeń, czy zupełnie nowa? Do rozważenia jest kilka kwestii.

- W obowiązujących rozporządzeniach została zdefiniowana struktura bazy danych w postaci schematu aplikacyjnego UML oraz standard wymiany danych w postaci schematu aplikacyjnego GML (jest on zapisywany jako zbiory z rozszerzeniem XSD). Ale nigdzie nie jest napisane, że struktura bazy danych musi być zgodna z UML z rozporządzenia. Natomiast dane z aplikacji muszą „wychodzić” (możliwość eksportu danych) i dane muszą do niej „wchodzić” (możliwość importu danych) w postaci plików GML, których struktura jest zgodna ze zdefiniowaną w rozporządzeniach.

- Trzeba jednak mieć świadomość, że im bardziej struktura bazy danych będzie różna od modelu UML, tym szybciej będą rosły kłopoty przy nowelizacjach rozporządzeń, a takie nowelizacje nastąpią.

- „Stare” dane (czyli te w dotychczasowej strukturze, ale już dzisiaj nieak-

tualne) muszą zostać zachowane i powinniśmy mieć do nich dostęp w wszelki wypadek (np. w razie sprawy w sądzie). Wypadałoby więc albo zachować dotychczasowe oprogramowanie na 1-2 komputerach albo utworzyć model struktury bazy danych na tyle elastyczny, by wczytał te „stare” dane. Także te, które mają braki (a braki mogą mieć też dane aktualne – jak było wyżej wspomniane), a teraz nie ma możliwości ani organizacyjnych, ani finansowych, by te braki uzupełnić. Aplikacja musiałaby więc zostać „wyposazona” w bazę danych o takiej „elastycznej” strukturze. Braki, zwłaszcza dla danych aktualnych, byłyby w takiej sytuacji sukcesywnie uzupełniane w trakcie normalnego funkcjonowania bazy. Skutkiem wybrania tej drogi będzie „druga” baza z danymi niekompletnymi obok „pierwszej” bazy z danymi pełnymi. Takie rozdwojenie jest kłopotliwe, wymaga pamiętania o konieczności uzupełniania braków, więc najlepszym wyjściem jest usunięcie braków podczas procesu transformacji lub tworzenia nowych obiektów. Innym rozwiązaniem jest nic nie robić i poczekać. Tylko na co? A poza tym Unia wymaga, by część danych w pewnym określonym czasie została opublikowana. Gdy ten okres minie i nic nie będzie opublikowane, posyła się kary.

- Jako że mamy już XXI wiek, można przy okazji zmian zastanowić się nad

REKLAMA

Brak korekt z ASG-Eupos? Brak zasięgu sieci komórkowej? Zbyt mała ilość widocznych satelitów?
To już historia! Wypróbuj nowoczesne odbiorniki GNSS

NAVCOM
A John Deere Company
z możliwością odbioru sygnału z globalnej sieci...

STARFIRE!

ART-GEO
z nami mierzyć zawozi i wzroście

tel. 531 70 00 70
www.GNSS.net.pl

aplikacją, która „obejmie” swoim działaniem wszystkie „geodezyjne” aspekty, czyli np. dla powiatu nad jedną wspólną aplikacją dla danych GESUT, EGiB, BDOT500, osnów szczegółowych, adresów itp. Chyba czas najwyższy zacząć odchodzić od zbiorów aplikacji: jedna dla EGiB, inna dla osnów, a jeszcze inna dla mapy zasadniczej. Operowanie jedną aplikacją jest znacznie wygodniejsze.

• Funkcjonalności aplikacji. W przyszłości dane z ośrodka do geodetów oraz od geodetów do ośrodka będą przesyłane w postaci „n” zbiorów GML. Ta liczba „n” zależą będzie od liczby baz danych, do których będą należeć obiekty wchodzące w skład opracowania. Założymy, że dana praca będzie dotyczyła danych GESUT i BDOT500. W związku z tym geodeta powinien otrzymać z ośrodka trzy osobne pliki GML (trzy, ponieważ trzecim będą dane dotyczące osnowy). Także trzy pliki powinny wrócić do ośrodka. Tak sytuacja wygląda teraz. Oczywiście można było tak zbudować modele UML, by wydawany był jeden plik GML ze wszystkimi danymi. Ale wcześniej musiałyby zostać utworzone wszystkie bazy danych. A to chyba można na razie zaliczyć do sfery marzeń. Dlatego trzeba czekać, aż powstaną wszystkie bazy i potem ewentualnie nowelizować rozporządzenia.

• Przed uaktualnieniem baz danych na podstawie dostarczonych zbiorów GML wypadałoby przeprowadzić kontrolę nowych danych. Standardowo odbywa się kontrola formalnej zgodności plików GML ze standardem GML oraz zgodności ze strukturą zdefiniowaną w plikach XSD. W zakresie poprawności formalnej wykonują ją „walidatory”, czyli specjalne aplikacje. Jeśli wynik kontroli jest pozytywny, można zaktualizować dane, ale wersje, które właśnie są aktualizowane, nie będą usuwane, tylko przenoszone do archiwum. W archiwum (lub archiwach) powinny się znaleźć wszystkie wersje archiwalne wszystkich obiektów, które są lub kiedyś były zapisane w bazach danych w postaci wersji aktualnych. Danych w archiwum nikt nie może edytować, usuwać. Można je tylko kopiować.

Te wszystkie kwestie i zapewne kilkadziesiąt innych (jak np. zaznaczenie, które dane zostały wydane na zewnątrz i na jak długo, by nie wydawać ich powtórnie innej osobie czy firmie, tworzenie plików metadanych), o których tu nic nie napisałem, też należy wziąć pod uwagę przy podejmowaniu decyzji.

To, że rozporządzenia zostaną znówelizowane, jest pewne, pytanie tylko kiedy. Teraz w GUGiK są zbierane uwagi do nich (m.in. do modeli UML i GML). Przy nowelizacji część uwag zostanie uwzględnio-

na, więc zostaną zmienione modele UML i GML. W związku z tym w ogłaszanych przetargach na aplikacje nie należy stawiać twardego warunku, że nowa struktura ma być zgodna z tą z rozporządzeń, bo ona może zostać (a raczej na pewno będzie) zmieniona. Zmiany prawdopodobnie nie będą bardzo duże, ale będą. Dobrym pomysłem wydaje się np. wymaganie, by struktura bazy była zgodna z tą z rozporządzeń poza przypadkami, na które zgodzi się zamawiający. Bez takiej furtki (lub podobnej – w jej sformułowaniu przydadzą się prawnicy) zostaną przygotowane pliki GML, o które niekoniecznie zamawiającemu chodziło. Zwłaszcza że transformacja danych zajmie trochę czasu, przygotowanie zbiorów GML to kolejne „trochę czasu”, a w tym „trochę czasu” może mieć miejsce parę wydażeń, np. jakaś nowelizacja.

Nie należy też ogłaszać przetargów na transformację danych do nowej struktury (jeśli to możliwe) przed opracowaniem aplikacji. Jeśli nie ma aplikacji, to nie ma możliwości przeprowadzenia kontroli merytorycznej danych w utworzonych plikach GML. Do tego jest potrzebna aplikacja, do której można wczytać dane z plików GML, wyświetlić te dane na ekranie, skontrolować, a jeśli są błędy, oddać do poprawy. Bez aplikacji można tylko pliki GML sprawdzić „walidatorami” i odłożyć na półkę, by czekały, aż przyjdzie „świeższa” przyszłość, w której wszystko będziemy mieli i wszystko będzie działać.

W rozporządzeniach zostały opublikowane schematy UML, które definiują dane, jakie muszą być gromadzone. Ale jeśli w jakimś powiecie, mieście czy województwie rada powiatu czy samorząd uzna, że to jest za mało do prowadzenia bieżącej działalności, można te modele rozszerzyć. To rozszerzenie powinno jednak mieć solidne podstawy merytoryczne, ponieważ żaden urząd nie może sobie gromadzić danych, jakie chce, a tylko te, na które zezwala lub nakazuje prawo.

I po tych wszystkich rozważaniach czas na podjęcie decyzji: Czy chcemy zupełnie nową aplikację, czy też np. po zobowiązaniu się twórców dotychczasowego oprogramowania, że dostosują swój produkt do nowego prawa i będą tak czynić w przyszłości, decydujemy się pozostać przy oprogramowaniu używanym obecnie, ale w nowszej wersji?

• 4. Na zakończenie

Uważam, że dopiero po tych wszystkich analizach można się zdecydować na transformację danych do nowej struktury. Czyli wtedy, gdy wiemy, jakie dane i w jakiej postaci posiadamy, gdy wiemy, czego oczekujemy od aplikacji, i gdy taką aplikację mamy.

Proces transformacji, kontroli, rozwiązywania wielu problemów, które ujawniają się po drodze, tworzenia plików GML zajmie sporo czasu. W tym czasie nie można „zamknąć” bazy na klucz i „otworzyć”, gdy wszystko zostanie już wykonane. Okres potrzebny na utworzenie danych w nowej strukturze należy liczyć raczej w miesiącach (kilkanaście), może latach (gdy mówimy np. o dużych miastach), a na pewno nie w tygodniach. Przez ten czas wiele obiektów zostanie zmienionych. Transformację należy więc powtórzyć dla tych danych, ale to już powinno potrwać chwilę w porównaniu z pierwszą transformacją – ścieżka będzie przetarta i dopracowana w szczegółach. I błędy w nowych lub zmienionych danych w zasadzie nie powinny występować.

Dr inż. Zenon Parzyński



Konkurs na artykuł o IIP

W ramach projektu „IIP dla każdego” zapraszamy do udziału w konkursie na najlepszy artykuł na temat wykorzystania oprogramowania Intergraph do budowy IIP w administracji publicznej. W konkursie może uczestniczyć każdy z Państwa niezależnie od tego, czy jest przedstawicielem wykonawstwa geodezyjnego, administracji publicznej, geoinformatykiem czy zwykłym obywatelem. Nagrodą jest publikacja najlepszego artykułu w GEODECIE oraz wyjazd na międzynarodową konferencję użytkowników oprogramowania Intergraph ufundowany przez patrona projektu. **Na artykuły czekamy do 15 lutego.**

Regulamin na GeoForum.pl w zakładce IIP dla każdego

Redakcja