

INFRASTRUKTURA DANYCH PRZESTRZENNYCH WROCŁAWSKIEGO SYSTEMU INFORMACJI PRZESTRZENNEJ

SPATIAL DATA INFRASTRUCTURE IN THE WROCŁAW SPATIAL INFORMATION SYSTEM

Artur Barcikowski

Geomatic Sp. z o.o.

Tomasz Strzelecki

Politechnika Wrocławska

Słowa kluczowe: hurtownia danych przestrzennych, architektura fizyczna, import, konwersja i integracja danych przestrzennych, metadane

Keywords: spatial data warehouse, physical architecture, data import, conversion and integration, metadata

Streszczenie

W artykule omówiono zagadnienia związane z budowanym obecnie dla Urzędu Miasta Wrocławia zintegrowanym systemem informacji przestrzennej. Przedstawione w artykule prace stanowią jeden z etapów projektu realizowanego w ramach grantu celowego KBN. Zamieszczono tutaj koncepcję architektury fizycznej hurtowni danych przestrzennych omawiając jej strukturę oraz model funkcjonowania.

Wstęp

W ubiegłym roku Wrocław przystąpił do budowy miejskiego Systemu Informacji Przestrzennej. Prace realizowane są w ramach grantu celowego KBN nr 6T12 080 2001 C/5671.

Podstawowym celem realizacji projektu jest określenie procedur i mechanizmów wymiany danych przestrzennych wewnątrz Urzędu Miasta. Rezultatem prac będzie zdefiniowanie takiego rozwiązania, które umożliwi wymianę informacji pomiędzy poszczególnymi wydziałami UM oraz prowadzonymi przez nie systemami informatycznymi (budowanymi w różnym czasie, różnymi technologiami i dla różnych platform sprzętowo-programowych).

Dane przestrzenne zorganizowane zostaną w formie centralnej, wielotematycznej bazy danych (hurtowni danych), wykorzystującej mechanizmy RDBMS Oracle 9i z opcją Spatial Data Option, z możliwością ich prezentacji w narzędziach firm Bentley Systems i ESRI Corp.

Podstawą Wrocławskiego Systemu Informacji Przestrzennej będą dane funkcjonujące w Zarządzie Geodezji, Kartografii i Katastru Miejskiego – numeryczna mapa zasadnicza oraz graficzno-opisowa baza ewidencji gruntów i budynków. Z zasobem tym porównywane i weryfikowane będą wszystkie pozostałe bazy SIP wprowadzane do hurtowni danych.

Architektura fizyczna hurtowni danych przestrzennych

Struktura systemu hurtowni

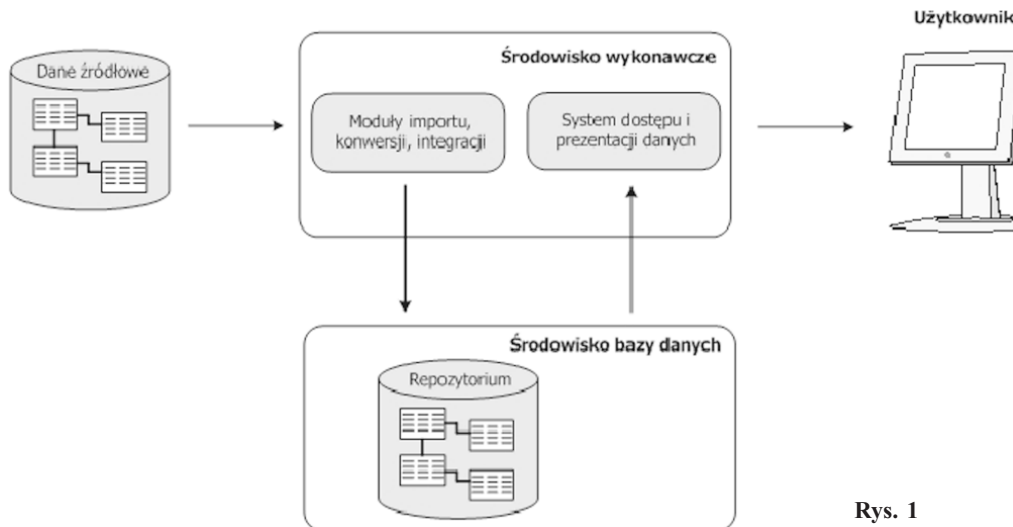
Głównym zadaniem hurtowni danych SIP jest skojarzenie w ramach jednego systemu danych o charakterze przestrzennym pochodzących z różnych źródeł (baz prowadzonych przez poszczególne wydziały Urzędu Miejskiego Wrocławia) oraz ich udostępnianie wszystkim zainteresowanym. Dlatego też struktura hurtowni danych zaprojektowana została tak, aby – z jednej strony – w maksymalny sposób umożliwić i ułatwić aplikacjom klienckim dostęp do danych oraz – z drugiej strony – zapewnić w pełni zautomatyzowany sposób importowania i integrowania danych modułom wykonawczym hurtowni. Założono, że moduły te oraz baza danych charakteryzować się będą następującymi własnościami:

- aplikacje klienckie będą miały możliwość odczytu danych, bez możliwości ich modyfikacji co powoduje, że baza danych będąca centrum hurtowni dla takiej aplikacji musi być zorientowana na realizację długich i złożonych odczytów,
- hurtownia danych nie będzie jednocześnie obsługiwała bardzo dużej liczby użytkowników, stąd nie będą występowały problemy związane z wielodostępem,
- hurtownia danych podlegać będzie okresowej aktualizacji (import dużej ilości danych) – proces importu oraz konwersja i integracja danych będą w praktyce jedynymi okresami wysokiej aktywności takiej bazy danych,
- w okresach pomiędzy aktualizacją danych zmienność danych będzie zerowa,
- o ile w systemach transakcyjnych bardzo ważny jest odpowiedni stopień normalizacji danych (redukuje to ilość redundancji i tym samym wpływa korzystnie na łatwość implementacji oraz wydajność modyfikacji danych) to podczas wprowadzania do hurtowni nowych danych, bardzo często będą one także podlegały procesowi odwrotnemu, tj. dodatkowo składowane będą w postaci zagregowanej, aby zminimalizować ilość niezbędnych złączeń i sortowań.

Określając architekturę fizyczną systemu przyjęto także, iż jego rolą jest dostarczanie danych podstawowych, a nie tworzenie struktury danych baz dziedzinowych (baz funkcjonujących w poszczególnych wydziałach UM). Taka baza danych charakteryzować się musi – w przeciwieństwie do znormalizowanej struktury bazy danych do przetwarzania transakcji – krótkimi czasami odpowiedzi na zapytania oraz strukturą, która jest klarowna zarówno dla administratorów, jak i użytkowników – nawet takich, którzy nie mają doświadczenia w pracy ze strukturami baz danych.

W swojej podstawowej strukturze, system hurtowni składać się będzie z dwóch środowisk: wykonawczego i bazy danych. Taki podział systemu przedstawiony jest na rysunku 1.

Centralne repozytorium zasilane będzie przez dane pochodzące z różnorodnych systemów dziedzinowych. Ten przepływ danych zostanie całkowicie zautomatyzowany, dzięki zastosowaniu specjalizowanych modułów importu, czyli komponentów przeznaczonych do pobiera-



Rys. 1

nia danych i metadanych płynących z określonych źródeł. W oparciu o zmiany zachodzące w systemach źródłowych, system hurtowni umożliwi aktualizowanie zgromadzonych w repozytorium danych, a także ułatwi tworzenie metadanych.

Na metadane składać się będzie m.in. lista baz danych, którymi zasilana jest hurtownia, struktury tych baz, reguły przesyłania danych z ich źródeł do hurtowni, zbiory raportów i zapytań, indeksy, reguły dzielenia obiektów hurtowni na partycje, reguły pobierania, przekształcania i korygowania danych źródłowych, uprawnienia użytkowników do danych zgromadzonych w hurtowni, sposób projekcji (odwzorowanie geograficzne) danych, itp. Aby zapewnić jakość i kompletność metadanych znajdujących się w repozytorium, system hurtowni będzie zawierał narzędzia sprawdzające poprawność danych. Dzięki nim możliwe stanie się zapewnienie spójności systemu, tworzonego z wielu baz dziedzinowych.

W skład środowiska wykonawczego wejdą narzędzia wspomagające proces importu, konwersji i integracji baz dziedzinowych oraz dostępu i prezentacji danych.

W środowisku bazy danych znajdują się tablice tymczasowe, które są niezbędne do poprawnego działania systemu oraz tablice systemowe hurtowni dla potrzeb repozytorium danych i metadanych, w których gromadzone będą informacje o atrybutach i powiązaniach wszystkich obiektów występujących w zasobie. Przewidziana jest trójwarstwowa architektura fizyczna bazy danych, obejmująca następujące wyodrębnione obszary:

- obszar roboczy, zawierający struktury niezbędne podczas procesów importu, konwersji i integracji danych z różnych źródeł,
- repozytorium, obejmujące zintegrowane zasoby danych i metadanych,
- obszar syntezy i agregacji, zawierający dane zagregowane, niezbędne do tworzenia wielowymiarowych przekrojowych raportów i sprawozdań oraz różnorodnych analiz, w tym wykonywanych na życzenie, zgodnie z pojawiającymi się potrzebami.

Takie zróżnicowanie zawartości i struktury bazy danych systemu hurtowni podyktowane jest przede wszystkim koniecznością włączenia do zasobu hurtowni jak największej ilości informacji oraz zapewnienia prostego i niezawodnego mechanizmu dostępu do danych.

Należy mieć świadomość, że przedstawiony model fizyczny nie jest jedynym rozwiązaniem dla hurtowni. Przedstawiona struktura bazy danych nie jest strukturą zamkniętą, tak więc możliwa będzie jej późniejsza rozbudowa o nowe obiekty oraz o nowe kategorie. Zaproponowany dla UM model centralnej bazy danych wynika ze szczegółowych rozważań na temat korzyści, jakie ma przynieść środowisko hurtowni danych, a nie z wiary w konkretne rozwiązania techniczne. Projekt ten nie stanowi sztywnej definicji architektury, lecz raczej spojrzenie na usługi i struktury dostępne w hurtowni oraz wykorzystywane zasoby z punktu widzenia wykonywanych procesów.

Środowisko wykonawcze

Spoiwem całego systemu hurtowni będą moduły importu, konwersji i integracji oraz dostępu do danych. Realizować one będą następujące funkcje:

- kwalifikacji i rejestracji zasobu danych,
- weryfikacji, konwersji i integracji danych w obszarze hurtowni,
- udostępniania jednolitego mechanizmu prezentacji danych i analiz,
- utrzymywania wewnętrznych standardów i reguł.

Dzięki tym specjalizowanym modułom, system hurtowni będzie potrafił korzystać z rozmaitych źródeł danych. Każdy z modułów zarządzać będzie danymi z odpowiedniej dziedziny (baza adresowa, mapa zasadnicza, ewidencja gruntów i budynków, itp.). System będzie miał możliwość swobodnego wyboru i łączenia ze sobą źródeł danych wraz z podglądem ich zawartości – jednak bez możliwości ingerencji w ich strukturę i zawartość. System w możliwie najbardziej zautomatyzowany sposób wykonywać będzie wszystkie niezbędne operacje mające na celu scalanie danych. Należą do nich: wstępna analiza danych (pod kątem struktury, typów, długości, nazewnictwa, itp.), lokalizacja wspólnych danych mogących dalej podlegać wspólnej analizie (nie wszystkie źródłowe bazy danych muszą zawierać identyczny zestaw informacji), ewentualnie odrzucenie duplikatów umożliwiającą wiarygodną i rzetelną analizę danych w hurtowni. Ostatecznym wynikiem działania modułów będzie spójny obraz jak największej ilości informacji dostarczonych z systemów źródłowych, który będzie mógł służyć do dalszej analizy z punktu widzenia hurtowni danych.

W systemie możliwe będzie raportowanie procesu wczytywania danych, dzięki udostępnionym procedurom monitorującym, których zadaniem będzie zapisywanie odpowiednich informacji zarówno do tabel metadanych wykorzystywanych w środowisku wykonawczym, jak i do plików typu log. Na informacje te składać się będą dane o ilości rekordów, jakie zostały wybrane, zapisane, bądź zaktualizowane oraz komunikaty o ewentualnych błędach występujących podczas wczytywania danych. Informacje te można będzie odczytywać przy pomocy narzędzi systemu hurtowni (tabele), a także wszelkiego rodzaju edytorów tekstu (pliki log).

Dodatkowo, co niezwykle ważne w przypadku baz danych przestrzennych, moduły systemu umożliwią automatyczną kontrolę i korektę topologii danych przestrzennych. Kontrola topologii będzie mogła wykrywać i w określonych przypadkach (np. przy zadanej tolerancji) usuwać następujące błędy:

- obiekty o tych samych cechach pokrywające się („zdublowane”),
- obiekty o tych samych cechach i bardzo zbliżonej geometrii,
- brak zamknięcia elementów powierzchniowych (niedociągnięcia, przeciągnięcia),
- brak ciągłości elementów liniowych,
- brak węzłów na przecięciach.

Wszystkie te funkcje będą realizowane poprzez odpowiednie moduły i procedury zarządzania i utrzymania hurtowni danych. Należy tutaj podkreślić, że hurtownia będzie zorientowana przede wszystkim na jak najszersze udostępnianie danych. Dlatego każdy podmiot będzie mógł zgłosić swój zasób do hurtowni, bez konieczności dbania o dostosowanie się do wykorzystywanej przez nią struktury danych. Każdy kto będzie chciał, aby dane z jego zasobu zostały zintegrowane z danymi hurtowni danych, powinien jedynie doprowadzić do tego, aby dane te spełniały podstawowe warunki integracji (takie jak precyzyjny opis zawartości, spójność, wiarygodność). Nie musi przy tym wykonywać we własnym zakresie żadnych operacji czyszczących czy transformujących – zajmą się tym nowe, specyfikowane moduły importu hurtowni.

System dostępu i prezentacji danych, odpowiedzialny będzie za udostępnianie danych przeznaczonych dla ściśle określonego użytkownika (wydziału) i ich analizę. Umożliwił on będzie dostęp do danych za pośrednictwem wyspecjalizowanych narzędzi analitycznych, pozwalających na generowanie raportów i prowadzenie różnorodnych, przekrojowych analiz. W systemie tym zastosowane będą rozwiązania informatyczne dedykowane dla systemów rozproszonych, wykorzystujące w zakresie komunikacji sieć lokalną (Intranet) i publiczną (Internet).

System dostępu i prezentacji danych oparty zostanie o wielowarstwową architekturę złożoną z:

- warstwy systemu zarządzania bazą danych,
- warstwy serwerów realizującej wymagane funkcje systemu,
- warstwy prezentacji danych.

Zastosowanie architektury wielowarstwowej da ogromną elastyczność zarówno podczas budowy systemu, jak i jego późniejszej rozbudowy. Z punktu widzenia celów długoterminowych drugi aspekt – rozbudowa systemu – jest w tym przypadku szczególnie ważny.

Środowisko bazy danych

Obszar roboczy. Obszar roboczy hurtowni to obszar wspomagający trzy podstawowe funkcje systemu hurtowni – import, konwersję i integrację. Funkcje te składają się na proces ładowania danych do hurtowni, który zwykle wymaga:

- zdefiniowania rodzaju baz źródłowych,
- zdefiniowania tabel źródłowych, z których będą pobierane dane,
- określenia odwzorowań źródło – repozytorium, stanowiących logiczne powiązania pomiędzy zdefiniowanymi tabelami źródłowymi a tabelami repozytorium hurtowni danych,
- opisanie sposobu filtrowania i przekształcania danych źródłowych,
- określenia metod integracji danych,
- określenia sposobu odświeżania danych w hurtowni danych odzwierciedlającego zmiany zachodzące w danych źródłowych.

Obszar roboczy pozwoli także na tworzenie wszechstronnych mechanizmów przepływu informacji z systemów dziedzinowych poprzez moduły importu (kontrolujące m.in. poprawność zapisu informacji) aż do tabel tymczasowych, wykorzystywanych do zasilania struktur repozytorium.

Podzielony on zostanie na dwie odrębne struktury danych:

- **obszar baz źródłowych**, zawierający struktury odpowiadające podstawowym bazom systemów dziedzinowych, przeznaczonych do wspierania wykonywanych przez wydziały UM zadań,
- **obszar tymczasowy**, obejmujący struktury niezbędne podczas procesów konwersji i integracji danych z różnych źródeł.

Obszar baz źródłowych będzie służył przechowywaniu zasobów stanowiących oddzielne bazy danych pochodzące z różnych systemów dziedzinowych w strukturze jednej bazy danych. Będzie on przechowywał pozyskane dane opisowe nie zmieniając ich struktury, natomiast dane graficzne zostaną transformowane na format Oracle Spatial. Taka struktura danych umożliwi prosty dostęp do baz źródłowych modułom konwersji.

Po odpowiednim załadowaniu danych źródłowych do obszaru baz źródłowych, następować będzie proces odwzorowania i konwersji ich na dane docelowe. System hurtowni udostępni moduły, przy pomocy których możliwe będzie odwzorowanie danych źródłowych na dane docelowe. Podczas odwzorowania brane będą pod uwagę tabele źródłowe i docelowe, a także wszystkie operacje mające wpływ na proces – generowanie brakujących wartości kluczowych, agregowanie danych, transformacje danych, usuwanie rekordów na podstawie zdefiniowanych filtrów, sortowanie danych, łączenie rekordów, itp. Na najniższym poziomie odwzorowania następować będzie kojarzenie kolumn tabel źródłowych z kolumnami tabel docelowych z wykorzystaniem logiki odpowiedzialnej za przekształcanie danych, przy czym niektóre atrybuty tabel źródłowych i docelowych odwzorowywane będą automatycznie. Dane będące końcowym wynikiem tych transformacji przechowywane będą w obszarze tymczasowym i uzupełniane dynamicznie podczas działania modułów importu, konwersji i integracji. Częścią tego procesu jest również łączenie wyczyszczonych i przekształconych danych w jeden zintegrowany zasób i przeniesienie go do repozytorium.

Jedną z głównych zalet obszaru roboczego hurtowni będzie możliwość odseparowania rzeczywistego repozytorium od źródłowych danych dziedzinowych i danych roboczych niezbędnych w trakcie wszystkich operacji, które są na nich wykonywane. Dzięki temu w przypadku zmiany struktury danych jednego z integrowanych systemów na inny (nowszy, różny funkcjonalnie), zmiany w strukturze danych hurtowni ograniczą się jedynie do niektórych struktur obszaru roboczego.

Repozytorium. Poza obszarem roboczym, system hurtowni obejmie także centralne repozytorium stanowiące główną część hurtowni danych. Repozytorium należy rozumieć jako tą część hurtowni, która odpowiada za składowanie, organizację, jakość i dostęp do właściwych danych. W proponowanym rozwiązaniu przyjęto, że repozytorium traktowane jest jako jeden zasób, do którego poprzez centralnie ulokowane moduły importu, konwersji i integracji wkładane będą obszerne wolumeny danych źródłowych.

Obszar syntezy i agregacji. Głównym zadaniem obszaru syntezy i agregacji będzie zmiana struktury zintegrowanych danych repozytorium do postaci, która lepiej nadaje się do prowadzenia przekrojowych analiz. W tym celu dane będą odpowiednio agregowane oraz zapisywane w wyspecjalizowanych strukturach obszaru w czasie procesu integracji.

Obszar ten będzie niezbędny, gdyż struktura baz dziedzinowych nie pozwala na efektywne generowanie raportów i analiz przekrojowych. Na przykład stworzenie zestawień opartych na danych wszystkich lokali komunalnych – w podziale na ich rodzaje i najemców, za ostatnie kilka lat – jest oczywiście możliwe, jednakże czasy odpowiedzi systemów mogą być bardzo

długie. Aby umożliwić interaktywne generowanie takich raportów przekrojowych, tworzenie zestawień ad-hoc i prowadzenie analiz, konieczne jest przetworzenie danych do postaci zoptymalizowanej pod kątem tego typu czynności.

Takie przygotowanie i zapisanie danych zagregowanych pozwoli m.in. na poprawienie wydajności zapytań użytkowników i zmniejszenie ogólnego czasu pracy jednostki centralnej.

Wdrożenie systemu

Projekt zakończony zostanie wdrożeniem pilotażowym zaproponowanego rozwiązania (planowany termin – listopad 2003 roku). Na jego podstawie można będzie określić przydatność zastosowanych rozwiązań technicznych, zweryfikować przewidywany nakład pracy nad rozbudową tematyczną i funkcjonalną hurtowni, wielkość niezbędnych zasobów ludzkich oraz nakładów finansowych, a także ocenić faktyczną przydatność posiadanych obecnie baz danych i ich kompletność.

Pilotaż odbędzie się w czterech wybranych jednostkach UM:

- Wydziale Mienia i Geodezji
- Wydziale Architektury i Budownictwa
- Wydziale Środowiska i Rolnictwa
- Zarządzie Zasobu Komunalnego.

Etap pilotowy dotyczy będzie ściśle określonego obszaru funkcjonalnego systemu i obejmować będzie następujące grupy tematyczne:

- bazę adresową
- ewidencję gruntów, budynków i lokali komunalnych
- mapę zasadniczą zawierającą wybrane elementy działów:
 - grunty
 - budynki
 - zagospodarowanie
 - komunikacja
 - rzeźba terenu i ogólnogeograficzne
- miejscowy plan zagospodarowania przestrzennego
- decyzje planistyczno-inwestycyjne.

Ustalając zakres informacyjny hurtowni danych przyjęto założenie, że jej rolą – na tym etapie – będzie dostarczanie danych najczęściej wykorzystywanych przez poszczególne wydziały UM. Pewne grupy tematyczne informacji przestrzennych zostały potraktowane bardziej szczegółowo z powodu ich szerszego wykorzystania. Chodziło tu o takie wyważenie zakresu tematyki projektowanej bazy danych, które zapewni jak największą uniwersalność systemu, tj. rozbudowanie tych informacji, które znajdują największą ilość użytkowników. Ważnym elementem przyjętej strategii było uwzględnienie istniejących lub obecnie tworzonych baz dziedzinowych, które będą zasilać hurtownię.

Literatura

- Heywood, I., Cornelius, S., Carver, S. *Introduction to Geographical Information Systems*, Prentice Hall, 1998
- Hohl P., *GIS Data Conversion: Strategies, Techniques, Management*, OnWord Press, 1998
- Plewe B., *GIS online, information retrieval, mapping, and the internet*, OnWord Press, 1997
- Poe V., Klauer P., Brobst S., *Tworzenie hurtowni danych*, Wydawnictwo Naukowo Techniczne, Warszawa, 2000
- Roszkowski J., *Analiza i projektowanie strukturalne*, Wydawnictwo Helion, 2000

Summary

This paper presents major aspects of currently developed integrated Spatial Information System for the Municipality of Wrocław. The tasks presented in the paper make up a stage of the project carried out within the framework of a KBN (Committee for Scientific Research) grant. The concept of the physical architecture of a spatial data warehouse is discussed, concentrating on its structure and functional model.

Artur Barcikowski
Geomatic Sp. z o.o., 53-439 Wrocław,
ul. Grabiszyńska 151
tel. (071) 361 44 11, fax (071) 361 44 15
e-mail: abarcikowski@geomatic.biz

Tomasz Strzelecki
Politechnika Wrocławska,
50-370 Wrocław, Wybrzeże Wyspiańskiego 27
tel. (71) 320 23 15,
e-mail: tomasz@strzelecki.net.pl