

PRÓBA AUTOMATYZACJI PROCESU GENERALIZACJI WYBRANYCH ELEMENTÓW BAZY DANYCH OGÓLNOGEOGRAFICZNYCH

AN ATTEMPT TO AUTOMATE THE GENERALIZATION PROCESS OF SELECTED COMPONENTS OF GENERAL GEOGRAPHIC DATABASE

Izabela Chybicka

Katedra Kartografii, Wydział Geografii i Studiów Regionalnych, Uniwersytet Warszawski

Słowa kluczowe: generalizacja danych przestrzennych, Baza Danych Ogólnogeograficznych
Key words: generalization of a spatial data, General Geographic Database

Cel i zakres opracowania

Celem opracowania jest przypisanie odpowiedniej ilości informacji do danego poziomu szczegółowości, a w rezultacie próba opracowania podstaw metodycznych generalizacji Bazy Danych Ogólnogeograficznych (BDO). Przedstawiany temat jest niezwykle istotny z punktu widzenia budowy infrastruktury danych przestrzennych w naszym kraju. Włączenie się Polski do inicjatywy INSPIRE wymaga dostarczenia społeczeństwu informacyjnemu szeroko rozumianej informacji przestrzennej na różnych poziomach szczegółowości.

Zakres opracowania obejmuje próbę automatyzacji generalizacji warstw tematycznych – sieci komunikacyjnej oraz osadnictwa – dla obszaru województwa dolnośląskiego. Badania te są kontynuacją wcześniejszych prac generalizacyjnych związanych z określeniem możliwości generalizacji baz danych przestrzennych (Chybicka, Iwaniak, Ostrowski 2004).

Badania dotyczące metodyki generalizacji map

Zajmowanie się zagadnieniem generalizacji danych przestrzennych w systemach GIS wymaga wcześniejszego zdefiniowania tego problemu. Przytoczyć można definicję zaproponowaną przez Liqiu Meng (1998), gdyż bierze ona pod uwagę generalizację w systemach GIS. Według Meng generalizacja jest procesem poznawczym, którego celem jest zachowanie czytelności obrazu odpowiednio do przestrzeni obszaru wyświetlania, rozdzielczości urządzenia wyświetlającego, poziomu percepcji i wymagań aplikacyjnych odbiorcy (Meng, 1998).

Jednym z głównych założeń automatycznej generalizacji jest poznanie procesu generalizacji z punktu widzenia jego formalizacji, a więc zaproponowania sekwencji elementarnych kroków, które wykonane w środowisku komputerowym (GIS) pozwolą na uzyskanie danych przestrzennych uogólnionych, o zakładanym stopniu szczegółowości. Należy jednak założyć, że podjęte czynności (kroki generalizacyjne) będą na tyle obiektywne, iż wykonane przez dwóch kartografów niezależnie, doprowadzą do tego samego rezultatu.

Od blisko dwóch dekad pozyskiwanie i implementacja wiedzy kartograficznej w systemach komputerowych jest najczęściej pojawiającym się zagadnieniem w literaturze kartograficznej. Tradycyjne metody statystyczne w zagadnieniach związanych z generalizacją zostały wyparte przez systemy oparte o reguły (systemy ekspertowe), dominujące jako technika gromadzenia wiedzy, którą pozyskiwano z istniejącej dokumentacji oraz na podstawie wywiadów z ekspertami, czy na bazie istniejących przykładów. Wiedzę tę opisywano w postaci reguł: „jeśli...to...”. Słabym punktem tak skonstruowanych systemów był jednak brak precyzji reguł, a w rezultacie brak jednoznaczności. Wiedza pozyskiwana od różnych ekspertów różniła się. Powstała, więc koncepcja by „nauczyć” maszynę pozyskiwania wiedzy eksperckiej tzw. „wyższego poziomu” lub strategii rozwiązywania problemów poprzez przykłady i wiedzę podstawową, analogicznie jak w procesie uczenia się człowieka. Powstały specjalne algorytmy uczące, a systemy wzbogacone o nie nazywamy systemami eksploracji danych (Meng, 2005). Wspomnieć można tutaj chociażby o systemach sieci neuronowych, uczonych na wzorcach danych za pomocą wspomnianych algorytmów, by następnie wykonywać wyuczone czynności na innym materiale. Dodatkową zaletą systemów eksploracji danych w stosunku do metod statystycznych, przemawiających za stosowaniem ich w procesie generalizacji jest fakt, iż w procesie statystycznym w celu uzyskania zmiennych zależnych należy posiadać wszystkie zmienne wejściowe niezależne jako dane, natomiast w systemach eksploracyjnych dopuszczalna jest również sytuacja, w której bez znajomości wszystkich zmiennych niezależnych znajduje się optymalną przestrzeń rozwiązania. W literaturze znanych jest szereg technik eksploracji danych, takich jak: drzewo decyzyjne, logika rozmyta czy sztuczne sieci neuronowe. Ich zastosowanie w procesie automatyzacji generalizacji map małoskalowych, gdzie nie ma kompletnych instrukcji i poszukuje się pewnego optymalnego przedziału rozwiązań, mogłoby dać interesujące rezultaty.

Baza Danych Ogólnogeograficznych

Przedmiotem rozważanego procesu generalizacji jest Baza Danych Ogólnogeograficznych. Baza ta jest podstawowym komponentem Krajowego Systemu Informacji Geograficznej, zawierającym dane przestrzenne stanowiące odniesienie dla innych danych i obiektów. Umożliwia to identyfikację obiektów referencyjnych dotyczących pokrycia terenu i użytkowania ziemi. W skład BDO wchodzi następujące warstwy tematyczne:

- podział administracyjny,
- osadnictwo i obiekty antropogeniczne,
- hydrografia,
- rzeźba terenu,
- transport,
- pokrycie terenu i użytkowanie ziemi,
- obszary chronione i zamknięte,
- nazwy geograficzne.

Opracowanie dotyczy warstwy sieci komunikacyjnej, w skład której wchodzi kolejki i drogi oraz warstwy osadnictwa, a więc miejscowości – prezentowane za pomocą sygnatur i kontury zabudowy, dla większych miejscowości.

Narzędzia generalizacji

Do przygotowania eksperymentu zastosowano komercyjne oprogramowanie firmy Intergraph: GeoMedia służące do wykonywania zaawansowanych analiz przestrzennych oraz DynaGEN – specjalistyczny system do wspomaganie procesu generalizacji map.

W realizacji eksperymentu wykorzystano model McMastera i Shea. Bardziej obszerny opis tego modelu oraz funkcjonowania postulowanych w nim operatorów generalizacji i realizujących je algorytmów zawiera praca Iwaniak, Paluszyński, Żyszkowska (1998). Pojęcie operatora generalizacji zostało zdefiniowane jako elementarne przekształcenie (transformacja) mapy, które można wyrazić formułą matematyczną, bądź jednoznacznym opisem procedury (algorytmu). Takie przekształcenie możemy nazwać krokiem generalizacji.

Proces generalizacji komputerowej można określić jako sekwencję takich przekształceń, z podaniem wartości odpowiednich parametrów. Ta sekwencja i parametry muszą być dobrane w sposób zachowujący pewne warunki i związki pomiędzy generalizowanymi obiektami.

W trakcie procesu generalizacji, w zależności od rodzaju generalizowanego obiektu, kartograf ma do dyspozycji szereg operatorów, algorytmów i parametrów generalizacji (operator upraszczania *Simplify*, wygładzania *Smoothing*, agregacji *Aggregation*, zmiany sposobu przedstawiania obiektów *Collapse*, rozciągania granic *Boundary Extend*, wyboru obiektów reprezentatywnych *Typify*, prostowania kątów *Square*, łączenia obiektów *Merge*).

Generalizacja sieci komunikacyjnej i osadnictwa

Operacje wchodzące w skład procesu generalizacji można klasyfikować w różny sposób. Autorka zgadza się z koncepcją podziału tego procesu na generalizację modelu danych oraz generalizację kartograficzną proponowaną przez Bell, Neuffer, Woodsford (2004).

Generalizacja modelu pozwala na redukcję ilości danych w stosunku do zakładanego poziomu szczegółowości. Obejmuje ona następujące czynności:

- selekcję całych klas obiektów,
- wybór podzbioru obiektów z danej klasy na podstawie warunków atrybutowych i przestrzennych,
- zmiany typu geometrii obiektu (sposobu ujęcia, a co za tym idzie metody prezentacji obiektu),
- upraszczanie geometrii.

Rolą generalizacji kartograficznej jako kolejnego etapu po generalizacji modelu danych jest osiągnięcie optymalnej czytelności mapy w danej skali, z punktu widzenia celu jej opracowywania. Generalizacja kartograficzna obejmuje następujące procesy:

- zastosowanie odpowiedniej symboliki dla danych,
- przesuwanie obiektów, agregacja, zmiana wymiaru obiektów.

W artykule przedstawiono pierwszy etap procesu generalizacji, związany z generalizacją modelu danych. Dobór treści dla wizualizacji w poszczególnych stopniach szczegółowości

wykonano na podstawie analizy istniejących map ogólnogeograficznych oraz wywiadów z ekspertami w zakresie generalizacji. Dobór treści do wizualizacji w każdej ze skal (1: 500 000, 1: 1 000 000 oraz 1:4 000 000) obejmował wykonanie odpowiednich analiz przestrzennych i atrybutowych w systemie GeoMedia. Operacje związane z upraszczaniem i wygładzaniem obiektów wykonano w systemie DynaGEN. Proces ten obejmował wybór dróg, uproszczenie i wygładzenie ich przebiegu oraz agregację i uproszczenie konturów zabudowy.

Generalizacja sieci komunikacyjnej oraz osadnictwa dla poszczególnych poziomów skalowych

Poniżej opisano podstawowe kroki generalizacyjne wchodzące w skład generalizacji sieci komunikacyjnej oraz osadnictwa dla poszczególnych poziomów skalowych. Proces generalizacji sieci komunikacyjnej poprzedziło wstępne przygotowanie danych. Polegało ono na łączeniu mniejszych segmentów dróg (powstałych wskutek wektoryzacji) w większe, ciągle obiekty, wykorzystywane do procesu generalizacji interaktywnej. Czynność tę wykonano w środowisku DynaGEN przy pomocy operatora łączenia (*Feature blending, Merging*) elementów obiektów w struktury sieciowe. Kryterium łączenia obiektów stanowi taka sama wartość atrybutu związanego z numerem drogi międzynarodowej lub krajowej oraz jednostki zarządzającej drogą (województwo, powiat, gmina itp). Proces wykonany został w trybie automatycznym, na podstawie zidentyfikowania obiektów o punktach wspólnych oraz takich samych wartościach wspomnianych atrybutów. Kolejne kroki generalizacji dróg i kolei zostały wykonane w trybie interaktywnym. Przy generalizacji osadnictwa nie było potrzeby wstępnego przetwarzania danych, poszczególne kroki generalizacji wykonano w związku z tym wyłącznie w trybie interaktywnym.

Czynności generalizacyjne dla wizualizacji o stopniu szczegółowości odpowiadającym skali 1: 500 000:

- wybór miejscowości przedstawionych sygnaturą – pozostawiono miasta, będące siedzibami władz wojewódzkich, powiatowych oraz gminnych;
- wybór dróg – pozostawiono drogi krajowe, wojewódzkie oraz powiatowe;
- wybór kolei – odrzucono koleje wąskotorowe oraz nieczynne;
- wybór miejscowości przedstawionych konturem – pozostawiono kontury zabudowy o powierzchni większej niż 9 mm^2 w skali mapy oraz kontury miast będących siedzibami władz wojewódzkich i powiatowych;
- uproszczenie przebiegu dróg – czynność wykonano przy pomocy operatora upraszczania *Simplify*; algorytmu *Douglas*; wartość parametru tolerancji dobrano eksperymentalnie, wynosiła 0,15;
- wygładzenie przebiegu dróg – czynność wykonano przy pomocy operatora wygładzania *Smooth*; algorytmu *Simple Average*; wartość parametru *Look Ahead* = 3;
- uproszczenie przebiegu kolei – zastosowano operator *Simplify*; algorytm *Douglas*; wartość parametru tolerancji wynosiła 0,35;
- wygładzenie przebiegu kolei – czynność wykonano przy pomocy operatora wygładzania *Smooth*; algorytmu *Simple Average*; wartość parametru *Look Ahead* = 3;
- uproszczenie konturów zabudowy – wykorzystano operator: *Simplify*; algorytm *Area Preservation*; wartość parametru *Area Change Allowed* = 0,15;
- wygładzenie konturów zabudowy – zastosowano operator: *Smooth*; algorytm: *Simple Average*; wartość parametru *Look Ahead* = 1.

Czynności generalizacyjne dla wizualizacji o stopniu szczegółowości odpowiadającym skali 1:1000 000:

- wybór miejscowości przedstawionych sygnaturą – pozostawiono miasta, będące siedzibami władz wojewódzkich oraz powiatowych; zmieniono sposób ujęcia, a co za tym idzie metodę prezentacji obiektów z konturu na sygnaturę;
- wybór dróg – pozostawiono drogi krajowe i wojewódzkie;
- wybór miejscowości przedstawionych konturem – pozostawiono kontury zabudowy o powierzchni większej niż 9 mm² w skali mapy;
- uproszczenie przebiegu dróg – zastosowano operator *Simplify*; algorytm *Douglas*; wartość parametru tolerancji = 0,20;
- wygładzenie przebiegu dróg – czynność wykonano przy pomocy operatora *Smooth*; algorytmu *Simple Average*; wartość parametru *Look Ahead* = 3;
- uproszczenie przebiegu kolei – operator *Simplify*; algorytm *Douglas*; wartość parametru tolerancji = 0,35;
- wygładzenie przebiegu kolei – zastosowano operator *Smooth*; algorytm *Simple Average*; wartość parametru *Look Ahead* = 3;
- uproszczenie konturów zabudowy – operator *Simplify*; algorytm *Area Preservation*; wartość parametru *Area Change Allowed* = 0,25;
- wygładzenie konturów zabudowy – zastosowano operator *Smooth*; algorytm *Simple Average*; wartość parametru *Look Ahead* = 1.

Czynności generalizacyjne dla wizualizacji o stopniu szczegółowości odpowiadającym skali 1:4 000 000:

- wybór miejscowości przedstawionych sygnaturą – sygnaturą pokazano miejscowości prezentowane w postaci konturów w skali 1:1000 000; zmieniono metodę prezentacji obiektów z konturu na sygnaturę;
- wybór dróg – pokazano drogi krajowe;
- uproszczenie przebiegu dróg – zastosowano operator *Simplify*; algorytm *Douglas*; wartość parametru tolerancji przyjęto 0,30;
- uproszczenie przebiegu kolei – zastosowano operator *Simplify*; algorytm *Douglas*; wartość parametru tolerancji = 0,40;
- wygładzenie przebiegu kolei przy pomocy operatora wygładzania *Smooth*; algorytm *Simple Average*; wartość parametru *Look Ahead* = 3.

Wizualizacja BDO w założonych stopniach szczegółowości

Ostatecznej wizualizacji zgeneralizowanych danych dokonano w systemie GeoMedia. Efekty generalizacji i doboru treści dla poszczególnych poziomów szczegółowości widoczne są na rysunkach 1, 2, 3 i 4.

Podsumowanie

Przeprowadzony eksperyment pozwala wyciągnąć kilka istotnych wniosków. Z punktu widzenia dotychczasowych osiągnięć w zakresie generalizacji map małoskalowych, przy opracowywaniu map należałoby uwzględnić szereg innych kryteriów generalizacyjnych. Jednakże założeniem opracowania jest określenie możliwości generalizacji BDO, na podstawie zawartych w niej informacji oraz otrzymanie powtarzalnych, obiektywnych rozwiązań, dlatego też ograniczono się do zawartości samej bazy. Podczas wykonywania eksperymentu napotkano trudności w implementacji zasad generalizacji. Utrudnienia te związane były z brakiem wystarczająco precyzyjnych narzędzi analiz przestrzennych (na przykład w narzędziach identyfikujących w sposób jednoznaczny odcinki „ślepo-zakończony”), co należałoby w przyszłości udoskonalić przez opracowywanie własnych narzędzi dostosowanych do konkretnych przypadków graficznych.

W procesie generalizacji BDO należy niewątpliwie uwzględnić gęstość sieci osadniczej, co zostanie wzięte pod uwagę w dalszych etapach badawczych.

Generalizacja w skalach małych – przeglądowych jest procesem niezwykle subiektywnym, intuicyjnym. Decyzje w zakresie podstawowych czynności generalizacyjnych w znacznej mierze zależą od wiedzy i praktyki kartograficznej osoby wykonującej mapę. Ze względu na charakter tego procesu nie istnieją instrukcje redakcji czy generalizacji map przeglądowych w tak dokładnej i sformalizowanej formie jak dla skal większych, co znacznie utrudnia automatyzację procesu.

Specyficzny charakter procesu generalizacji map w tych skalach niejednokrotnie wymaga różnych rozwiązań tego samego problemu, w zależności od otoczenia i kontekstu obiektów.

Dążenia do formalizacji i automatyzacji procesu generalizacji danych przestrzennych koncentrują się na opracowaniu takich systemów, które posiadałyby niezbędną, a charakterystyczną wyłącznie dla człowieka, wiedzę i doświadczenie – niemalże zdolność abstrakcyjnego myślenia i podejmowania decyzji. Na pytanie czy jest to obecnie możliwe, można odpowiedzieć, że jeszcze nie jest. Jednakże wysiłki zmierzające do automatycznego pozyskiwania, a następnie wykorzystania wiedzy posiadanej przez człowieka w środowisku komputerowym są bardzo obiecujące i ten właśnie kierunek badań cieszy się dużym zainteresowaniem naukowców.

Literatura

- Bell M., Neuffer D., Woodsford P., 2004: Agent-based generalization – an update on progress, *Kartographische Nachrichten*, Vol 54, nr 4, s. 170-177.
- Chybicka I., Iwaniak A., Ostrowski W., 2004: Generalization of the Topographic Database to the Vector Map Level 2 – the components of the Polish National Geographic Information System. <http://ica.ign.fr/Leicester/paper/Chybicka-v2-ICAWorkshop.pdf>
- Chybicka I., Iwaniak A., Ostrowski W., Paluszyński W., 2004: Generalizacja danych topograficznych o szczegółowości 1:10 000 do skali 1:50 000. Cz. II, *Polski Przegląd Kartograficzny*, T. 36, nr 4, s. 266-273.
- Iwaniak A., Paluszyński W., 2003: Implementation of a knowledge database for the generalization of topographic maps in GIS systems. [W:] International Cartographic Conference. Abstracts of papers, Durban, South Africa, s. 36.
- Iwaniak A., Paluszyński W., Żyszkowska W., 1998: Generalizacja map numerycznych – koncepcje i narzędzia cz.I., *Polski Przegląd Kartograficzny*, T. 30, nr 2, s. 79-86.

- Meng L., 1998: Cognitive modeling of cartographic generalization, Project report on Strategies on automatic generalization of geographic data – stage2, Hogskolan i Gavle, Sandviken, s. 1-31.
- Meng L., 2005: Spatial data mining – methods and case studies, Prezentacja podczas seminarium nt.: Zastosowanie sztucznej inteligencji w kartografii, Politechnika Warszawska.

Summary

One of the main assumptions of an automatic generalization is understanding of the process taking into account its formalization. Such an approach is based on a proposal of performing elementary steps in computer GIS environment in order to obtain generalized spatial data at particular detail level. However, it should be expected that the generalization steps are so objective that when carried out by two cartographers independently they bring common result.

As the National Geographic Information System and its components are developed in Poland (its components include: Topographic Database at 1:10 000 scale, Vector Map Level 2 at 1:50 000 scale, General Geographic Database at 1:250 000 scale), topicality of the generalization grows ever larger. Under such conditions there is a need for creating a uniform spatial database out of which maps at various scales and for various purposes could be generated.

The executed project is extremely important from the point of view of building spatial data infrastructure in our country. Having joined the INSPIRE initiative, Poland is required to provide the information society with well-understood spatial data collected at different resolution levels. However, the purpose of the project is to define and then to assign a particular portion of information to a particular resolution level and, consequently, to work out a methodology of generalization of the basic spatial database GGD.

The scope of the study covered carrying out generalization experiment concerning the generalization of thematic layers – transportation network and settlement for the area of the Lower Silesia Province. This study was a continuation of previous works concerning generalization possibilities of spatial databases (I. Chybicka, A. Iwaniak, W. Ostrowski, 2004).

The intention was to elaborate a multi-resolution/representation database as visualization of the General Geographic Database (GGD) at different resolution levels without permanent loss of information. The selection of information depends not on durable data withdrawal from a database but only on visualization of the generalized information adequately to resolution level. The visualization of the GGD was performed based on three levels: 1:500 000, 1:1000 000 and 1:4000 000.

The operations forming the generalization process may be classified in different ways. The author agrees with the concept of dividing the process into the data model generalization and the cartographic generalization proposed by M. Bell, D. Neuffer, P. Woodsford (2004).

The data model generalization makes it possible to reduce the number of data in relation to the assumed resolution level. It covers the following actions:

- selection of whole feature classes,*
- selection of object components from a particular feature class on the basis of attributes and spatial conditions,*
- change of object's geometry type (way of presentation and method of object's presentation),*
- geometry simplification.*

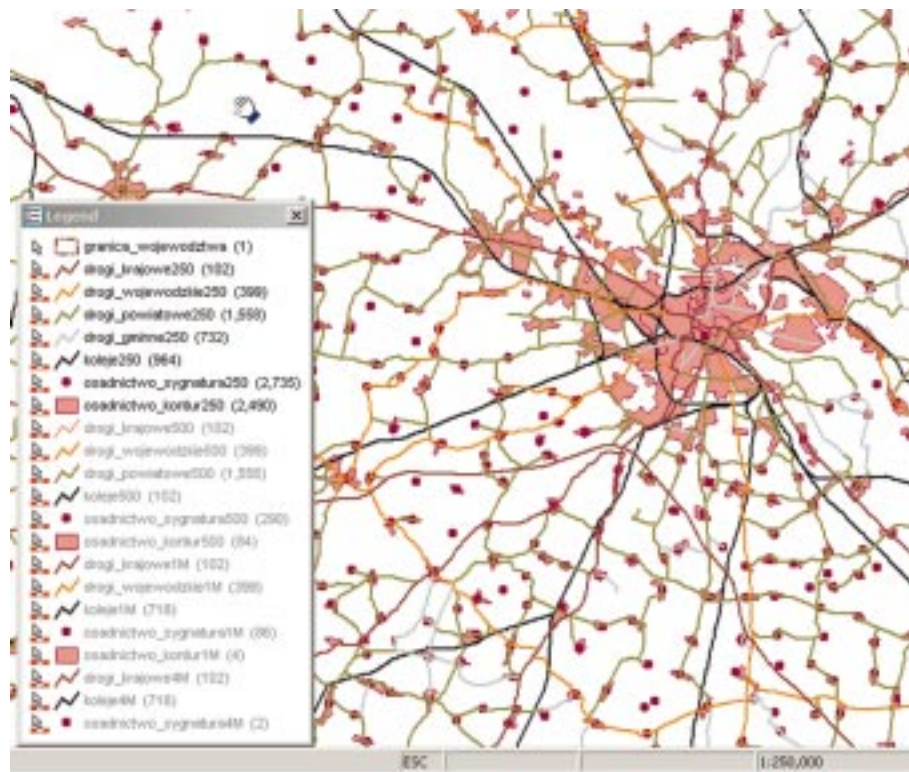
The role of the cartographic generalization as a stage following the data model generalization is to obtain optimal map legibility at a given scale. The cartographic generalization process consists of:

- application of proper data symbology,*
- shifting of objects; aggregation; changing of object's dimensions.*

The author focused on the first stage of the generalization process (connected with the data model generalization). The selection of the map content for visualization on particular resolution degrees was performed on the basis of analysis of existing geographic maps as well as interviews with experts in the field of generalization. Ordering the map content to the visualization performed for each of the

scales (1: 500 000, 1: 1000 000, 1: 4000 000) covered performing proper spatial and attribute analyses in the GeoMedia system. The operations connected with the simplification and objects' smoothing were performed in the DynaGEN application. This process encompassed simplification of routes and their smoothing as well as simplification and smoothing of buildings' contours.

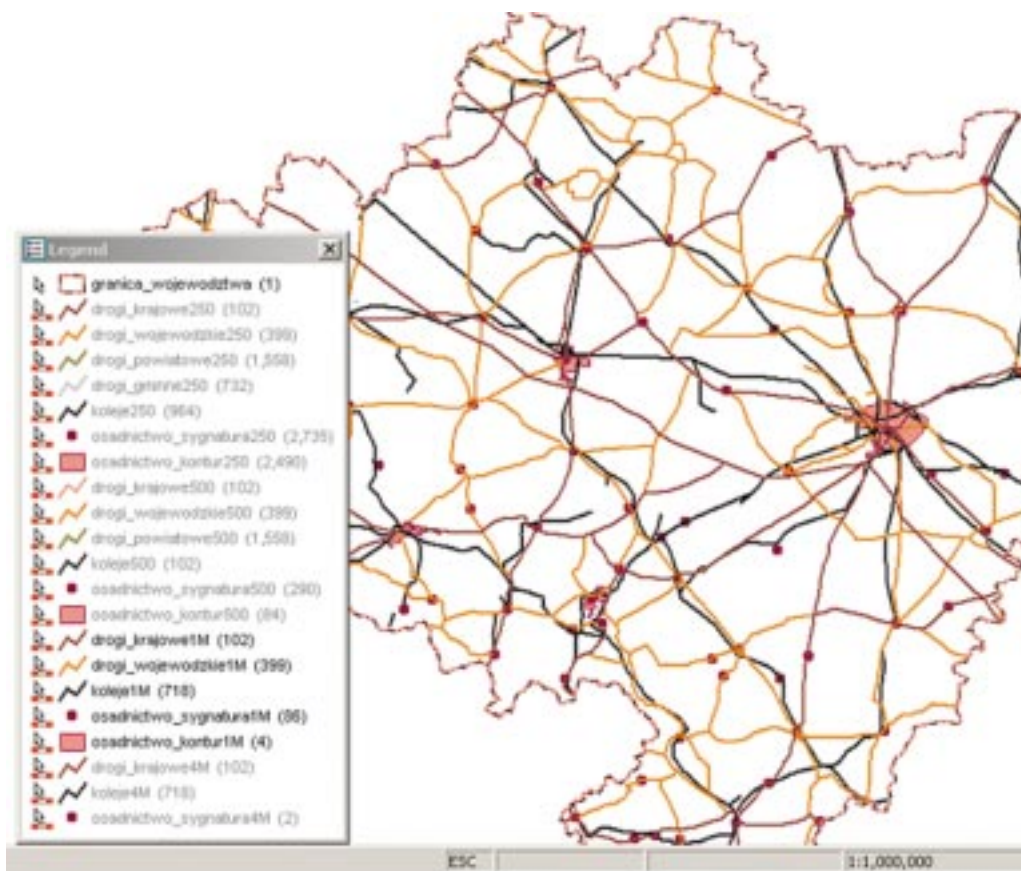
mgr inż. Izabela Chybicka
doktorantka UW
i.chybicka@uw.edu.pl
tel. (0-22) 552-15-11



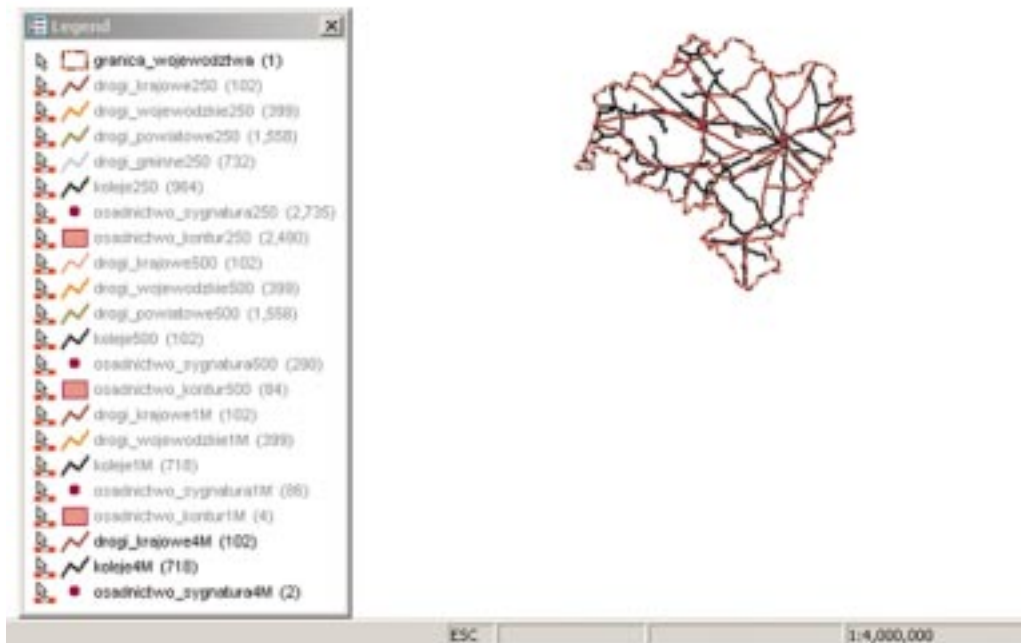
Rys. 1. Wizualizacja BDO w skali 1:250 000



Rys. 2. Wizualizacja BDO w skali 1:500 000



Rys. 3. Wizualizacja BDO w skali 1:1000 000



Rys. 4. Wizualizacja BDO w skali 1: 4 000 000