

## EWALUACJA METODYKI PROWADZENIA ANALIZ PRZESTRZENNYCH W WIELOREPREZENTACYJNEJ BAZIE DANYCH TOPOGRAFICZNYCH

### EVALUATION OF METHODOLOGY OF SPATIAL ANALYSIS IN MULTIREPRESENTATION DATABASE

**Robert Olszewski, Anna Fiedukowicz**

Politechnika Warszawska, Wydział Geodezji i Kartografii, Zakład Kartografii

**Słowa kluczowe: wielorozdzielcza baza danych, MRDB, SQL, system informacji geograficznej**  
Keywords: multiresolution database, MRDB, SQL, geographic information system

### Wprowadzenie

Jak zauważa Makowski (2005) *istotą kartografii jest modelowanie i obrazowanie czasoprzestrzennych struktur geoinformacyjnych*. Modelowanie informacji geograficznej wymaga zaś dostrzeżenia, zrozumienia i wreszcie sformalizowania wzajemnych zależności pomiędzy komponentami środowiska przyrodniczego i elementami sytuacyjnymi oraz wysokościowymi modelu topograficznego. Mająca wielowiekowe tradycje metodyka kartograficzna umożliwia – z jednej strony – opisywanie świata językiem mapy, z drugiej zaś – poznanie przestrzeni geograficznej poprzez analizę baz danych przestrzennych i zgromadzonych w nich informacji (Olszewski, 2013).

Osiągnięciem, tzw. orientacji poznawczej w kartografii, jest stworzenie w latach 60. XX wieku koncepcji modelu i modelowania kartograficznego. Redakcja mapy, a także współcześnie bazy danych geograficznych, została utożsamiona z procesem modelowania konkretnego stanu rzeczywistości lub szerzej, z opisem jego stanu oraz zmian zachodzących w czasie i przestrzeni geograficznej (Czerny, 1990; Grygorenko, 1984). Orientacja poznawcza traktuje kartografię jako naukę zajmującą się odwzorowaniem i badaniem zjawisk występujących w przestrzeni geograficznej, z punktu widzenia ich rozmieszczenia, właściwości, współzależności i zachodzących zmian z pomocą obrazów kartograficznych. Punktem wyjścia koncepcji teoriopoznawczej jest założenie, iż w procesie poznania mapa (lub baza danych przestrzennych) pośredniczy pomiędzy podmiotem poznającym a przedmiotem poznawanym.

Rozwijana od połowy XX w. koncepcja kartograficznej metody badań (KMB), oparta o to założenie, rozszerzyła rolę kartografii (Saliszczew, 1955, 1968; Berlant, 1973, 1978) z odtwórczej (przedstawiania rzeczywistości) do poznawczej (analizowania stanu istniejącego za pomocą mapy – modelu rzeczywistości i uzyskiwania wartości dodanej).

Model kartograficzny, rozumiany jako specyficzne odwzorowanie bytów rzeczywistych, z pominięciem nieistotnych cech i związków, służący ukazaniu charakterystycznych aspektów badanego fragmentu rzeczywistości w stopniu umożliwiającym jego poznanie (Weibel, 1991. 1995), jest zatem podstawą prowadzenia wszelkiego rodzaju analiz, wykorzystujących technologie GIS. Model ten, uwzględniający cel i przeznaczenie opracowania, może być przy tym abstrahowany na dowolnie zdefiniowanym poziomie uogólnienia pojęciowego.

Poznawcza rola kartografii, w dobie ekspansji baz danych przestrzennych i systemów GIS, wydaje się nawet większa niż w czasie kiedy świat modelowany był w postaci map analogowych (Gotlib, Iwaniak, Olszewski, 2007). Rosnące możliwości obliczeniowe komputerów oraz mnogość dostępnych danych, dają potencjał do prowadzenia badań, o których prekursorzy KMB mogli jedynie marzyć. Jak zauważa Olszewski (2013) *istnieje swoista analogia pomiędzy kartografią (lub ściślej mówiąc – kartograficzną metodą badań) a metodą badań statystycznych. By móc wiarygodnie wnioskować o populacji, prowadzi się badania na odpowiednio dobranej próbie reprezentatywnej. Przez zastosowanie adekwatnego dla celu opracowania modelu kartograficznego i jego analizę w systemach geoinformacyjnych, możemy wnioskować nie o układzie „kresk na mapie”, lecz o wzajemnym oddziaływaniu obiektów terenowych, o zależnościach komponentów środowiska przyrodniczego, o relacjach i układach przestrzennych.*

## Język SQL jako narzędzie prowadzenia analiz przestrzennych

Wraz ze sposobem modelowania przestrzeni geograficznej – mapy analogowe powszechnie zastępowane są bazami danych przestrzennych – zmieniają się także narzędzia analiz informacji geograficznej. Dane przestrzenne zapisane w formacie wektorowym analizowane są głównie poprzez użycie deklaratywnego języka zapytań SQL (ang. *Structured Query Language*). Opracowany w latach 70. XX w., początkowo jako narzędzie tworzenia i modyfikowania relacyjnych baz danych (Date, Darwen, 1997) SQL został później rozwinięty i we współczesnych implementacjach systemów informacji geograficznej (GIS) rozszerzony, m.in. o operatory przestrzenne, takie jak: *contain*, *within*, *intersects*, które pozwalają na uwzględnienie relacji geometrycznych i topologicznych między poszczególnymi obiektami i klasami obiektów. Większość użytkowników narzędzi geoinformacyjnych używa głównie zapytań o charakterze atrybutowym, jednak wynika to głównie z braku wiedzy i biegłości, nie zaś z ograniczeń technicznych. Obecnie, praktycznie w każdym narzędziu GIS, możliwa jest realizacja kwerendy SQL typu – *Wybierz drogi krajowe przecinające Wisłę. Jaka jest ich sumaryczna długość?* Na tak postawione, całkowicie jednoznaczne pytanie, istnieje jednak wiele odpowiedzi. Nie tylko algorytm definiujący operator przestrzenny „przecina” jest tu istotny. O wiele ważniejsze jest źródło danych przestrzennych, zależnie od dokładności i aktualności danych źródłowych, możliwe jest uzyskanie bardzo różnych odpowiedzi. Wykorzystanie danych topograficznych sprzed 24 marca 2012 r. skutkuje np. nieuwzględnieniem mostu Marii Skłodowskiej-Curie w Warszawie, co bardzo istotnie zmienia układ sieci drogowej w stolicy.

Użycie SQL pozwala więc na pośrednie poznawanie przestrzeni geograficznej, należy jednak pamiętać, że ogromny wpływ na uzyskane wyniki kwerendy ma dokładność geometryczna danych zgromadzonych w bazie danych przestrzennych oraz przyjęty sposób modelowania. Na wynik wpływa zarówno wybór modelowanych obiektów, sposób definiowania

poszczególnych klas obiektów, jak i przyjęty poziom szczegółowości (atrybutowej i geometrycznej).

Zapytanie SQL odniesione do danych topograficznych o dokładności geometrycznej, odpowiadającej mapom analogowym w skali 1:10 000 zwróci skrajnie odmienne wartości, np. liczby i sumarycznej długości dróg w danym powiecie niż analogiczna kwerenda wykorzystująca dane pochodzące z bazy danych ogólnogeograficznych o dokładności odpowiadającej mapom w skali 1:250 000. Interesujące jest jednakże pytanie: czy przy zachowaniu spójnego modelu pojęciowego bazy danych georeferencyjnych i zunifikowanej metodyki generalizacji informacji geograficznej, możliwe jest ekstrapolowanie wyników analiz przestrzennych na poszczególne poziomy uogólnienia pojęciowego i geometrycznego LoD (ang. *level of details*) bazy danych, czy też wyniki te są całkowicie nieprzewidywalne? Innymi słowy celem prowadzonych badań jest uzyskanie odpowiedzi na pytanie: na ile adekwatne jest wieloskalowe wnioskowanie o zjawiskach i procesach, mających lokalizację przestrzenną na podstawie zunifikowanych i monoskalowych danych źródłowych.

## Wielorozdzielcze bazy danych przestrzennych

Przez setki lat ludzie opisywali otaczającą ich przestrzeń geograficzną posługując się językiem kartografii – rysowali mapy, analizowali je i wykorzystywali do orientacji. Współcześnie językiem opisu przestrzeni geograficznej są przede wszystkim bazy danych przestrzennych oraz komputerowe systemy informacji geograficznej (ang. *Geographical Information System*, GIS). Umożliwiają one nie tylko gromadzenie, ale i szybkie analizowanie danych geograficznych i związanych z nimi atrybutów opisowych (Gotlib, Iwaniak, Olszewski, 2007).

Powstające spontanicznie, pod koniec ubiegłego wieku, systemy informacji geograficznej budowane były jako odrębne, z reguły hermetyczne rozwiązania informatyczne. Dane przestrzenne, odnoszące się do tego samego obszaru, były gromadzone wielokrotnie przez różne instytucje, w różnych narzędziach GIS, różnych formatach, z różną dokładnością i zapisywane w różnych strukturach baz danych. Powodowało to nie tylko redundancję, lecz także znacząco podnosiło koszty tworzenia systemów GIS i utrudniało współpracę instytucjonalną. Przełomowe znaczenie dla zmiany tego stanu miała budowa infrastruktur geoinformacyjnych o charakterze kontynentalnym (np. INSPIRE) i krajowym (IIP).

Jednym z projektów, realizowanych w Polsce niejako przy okazji wdrażania INSPIRE i implementującej jej idee do prawa polskiego ustawy o infrastrukturze informacji przestrzennej, jest GBDOT – georeferencyjna baza danych obiektów topograficznych. Realizacja tego projektu umożliwi do końca 2013 roku zgromadzenie dla obszaru całego kraju cyfrowych danych topograficznych, o dokładności geometrycznej i poziomie uogólnienia pojęciowego, odpowiadającym klasycznym mapom analogowym w skali 1:10 000. W ciągu następnych 2 lat na podstawie źródłowych danych BDOT10k zostanie opracowana pochodna baza danych BDOO, odpowiadająca mapom w skali 1:250 000 i mniejszym. Realizacja tego zamysłu jest nie tylko wielkim wyzwaniem organizacyjnym, finansowym i logistycznym, lecz stanowić będzie także istotne osiągnięcie intelektualne, gdyż docelowa baza danych referencyjnych będzie pierwszą w Polsce i jedną z pierwszych w Europie baz danych przestrzennych o charakterze wielorozdzielczym.

Wielorozdzielczą bazę danych – bazę typu MRDB (*multiresolution/multirepresentation data base*) można określić jako bazę danych przestrzennych umożliwiającą przechowywanie

reprezentacji rzeczywistych obiektów geograficznych, na różnym poziomie uogólnienia (dokładności, precyzji, skali lub rozdzielczości). Istotą tego podejścia jest zdefiniowanie relacji pomiędzy obiektami reprezentującymi w bazie MRDB rzeczywisty obiekt topograficzny na różnych poziomach generalizacyjnych. Korzystając z bazy MRDB uzyskujemy:

- możliwość wieloskalowej analizy danych przestrzennych,
- możliwość automatycznego zasilania (propagacji) bazy danych na wielu poziomach skalowych, na podstawie aktualizacji modelu podstawowego,
- automatyzację procesu zasilania danymi przestrzennymi systemów produkcji map na różnym poziomie skalowym.

Poprawnie utworzona baza danych topograficznych stanowi podstawowe źródło danych, dla dowolnie zdefiniowanych analiz przestrzennych i kartograficznych opracowań tematycznych. Struktura opracowywanej obecnie w Polsce bazy danych obiektów topograficznych (BDOT10k) i bazy danych obiektów ogólnogeograficznych (BDOO) nawiązuje do idei wielorozdzielczych baz danych (MRDB), zawierając zarówno dwa zintegrowane komponenty topograficzne, jak i szereg pochodnych modeli kartograficznych, odpowiadających mapom analogowym w skali od 1:10 000, aż do 1:1 000 000.

Zachowanie harmonizacji poszczególnych poziomów uogólnienia, ich spójności oraz powiązań między nimi, powinno być osiągane przez otrzymywanie danych o mniejszej szczegółowości w wyniku generalizacji danych źródłowych, nie zaś poprzez „łączenie” źródeł o odmiennym modelu pojęciowym (rys. 1). Takie postępowanie jest tym bardziej uzasadnione, że zapobiega wielokrotnemu pozyskiwaniu danych, co jest zgodne z zapisami dyrektywy INSPIRE. Spójność modeli pojęciowych na różnych poziomach szczegółowości jest również dużo łatwiejsza do zachowania w momencie, gdy są one wywodzone z siebie nawzajem, nie zaś tworzone niezależnie. Odpowiednio zaimplementowane narzędzia generalizacyjne pozwalają też zachować informację o obiektach bardziej szczegółowych, które odpowiadają obiektowi w innej skali.

Warto zauważyć, iż koncepcja wielorozdzielczej i trójpoziomowej bazy danych georeferencyjnych, wykorzystującej jako dane źródłowe moduły TBD, VMap L2 i BDO, była rozważana w Polsce już w ubiegłej dekadzie (Gotlib, Iwaniak Olszewski, 2007; Bac-Bronowicz, Olszewski, 2008; Olszewski, 2009; Bac-Bronowicz, Olszewski, 2009). Znaczące różnice modeli pojęciowych tych baz (rys. 1) oraz nowe uwarunkowania prawne (dyrektywa INSPIRE i ustawa o IIP) oraz organizacyjno-finansowe (projekt GBDOT) spowodowały, że pomysł ten zarzucono, jednak doświadczenia zebrane w trakcie prac nad tym projektem pomogły w dalszych pracach związanych z tworzeniem bazy typu MRDB w Polsce.

## **Analiza kartometryczna danych topograficznych o różnej szczegółowości**

Jak zauważono we wstępie do artykułu, poprawnie rozumiane modelowanie kartograficzne pozwala nie tylko na gromadzenie danych przestrzennych w bazie danych (i ich wizualizację w postaci map topograficznych i tematycznych), lecz także na poznawanie przestrzeni geograficznej w sposób pośredni, tj. poprzez analizę zgromadzonych danych.

Zdaniem autorów kluczowe znaczenie ma jednak odpowiedź na pytanie czy, i jeśli tak, to w jaki sposób poziom szczegółowości danych źródłowych wpływa na wynik przeprowadzonej analizy? Czy analizując dane BDOT10k, o poziomie dokładności geometrycznej i po-

ziomie uogólnienia pojęciowego, odpowiadającym mapom analogowym w skali 1:10 000 otrzymamy wyniki zbliżone do tych, które byłyby efektem analizy danych BDOO, o dokładności odpowiadającej skali 1:250 000? Na ile dokładność geometryczna i przyjęta metodyka generalizacji determinują wiedzę o przestrzeni geograficznej, którą czerpiemy – w sposób pośredni – z analizy danych topograficznych i/lub ogólnogeograficznych?

Pod pojęciem „analizy” można rozumieć zarówno elementarne zapytania do bazy danych przestrzennych, określenie współzależności i relacji topologicznych pomiędzy różnymi warstwami informacyjnymi, jak i konstruowanie złożonych modeli decyzyjnych (Gotlib, Iwaniak, Olszewski, 2007).

Analizy przestrzenne umożliwiają uzyskanie odpowiedzi na pytania typu: gdzie się znajduje? jaka jest zależność przestrzenna pomiędzy obiektami lub klasami obiektów? co zmieniło się od momentu? czy istnieje ogólna prawidłowość w rozkładzie danych przestrzennych? (Longley, 2006). Analiza danych georeferencyjnych umożliwia przekształcenie danych źródłowych w informację, będącą źródłem poznania otaczającej przestrzeni i podstawą wspomaganego decyzyjnego. Wyniki realizowanych analiz zależą nie tylko od przestrzennej lokalizacji danych źródłowych, lecz także od dokładności geometrycznej i poziomu uogólnienia (skali) informacji przestrzennych, wiarygodności i kompletności atrybutów opisowych, aktualności danych przestrzennych i opisowych, a nawet zastosowanych procedur obliczeniowych i narzędzi GIS.

Analiza danych przestrzennych może być realizowana zarówno *a priori*, jak i *a posteriori*. Pozwala to na objaśnienie historycznych procesów naturalnych i antropogenicznych, mających odniesienie przestrzenne, a także na budowanie scenariuszy prognostycznych, uwzględniających całe spektrum parametrów wejściowych. Tak zbudowany system umożliwia także testowanie hipotez i szacowanie potencjalnych skutków prognozowanych zmian.

Specyfika modelowania kartograficznego w wielorozdzielczej bazie danych typu MRDB sprawia, że konieczne wydaje się przetestowanie, jak istniejące narzędzia analiz (np. język SQL z rozszerzeniami do prowadzenia analiz przestrzennych) sprawdzają się dla danych wielorozdzielczych. Powszechnie stosowane implementacje SQL służą do analizy danych homogenicznych, pod względem poziomu szczegółowości. Analiza prowadzona na danych o charakterze wieloskalowym zależna jest od wielu czynników, z których do najistotniejszych należą:

- 1) zastosowany model pojęciowy np. definicje klas obiektów (rys. 2),
- 2) przyjęty poziom szczegółowości i uogólnienia pojęciowego (rys. 3),
- 3) wykorzystana metodyka generalizacji (np. algorytmy upraszczania kształtu – rys. 4)
- 4) aktualność wykorzystywanych danych – w zależności od momentu pomiaru/aktualności danych źródłowych analiza przestrzenna da inny efekt (rys. 6).

W przypadku większości baz danych przestrzennych, monoskalowe dane o jednym poziomie szczegółowości, determinują jednoznaczny wynik analizy przestrzennej. Ze względu na specyfikę baz wielorozdzielczych, istniejące narzędzia analityczne nie są jednak wystarczające, aby w pełni wykorzystać potencjał zgromadzonych danych. Wyniki realizowanych kwerend o charakterze przestrzennym mogą się istotnie różnić, w zależności od stopnia szczegółowości danych źródłowych. Dlatego też, zasadne może się okazać zaproponowanie nowych rozwiązań, które uwzględniłyby w pełni różne reprezentacje tych samych obiektów, wykorzystując np. powiązania między nimi, zapisywane w bazie w trakcie tworzenia wyższych stopni uogólnienia, w ramach procesu generalizacji.

## Metodyka przeprowadzonego eksperymentu

Do przeprowadzenia eksperymentu autorzy wykorzystali dane przestrzenne o jednorodnym pochodzeniu i spójnym modelu pojęciowym, a także jednakowej aktualności, pochodzące z zasobu geodezyjnego i kartograficznego i utożsamiane z bazą referencyjną BDOT10k (określane dalej jako TOPO10). Dysponując danymi dla obszarów testowych oraz ogólnymi wytycznymi w zakresie generalizacji informacji geograficznej, zdefiniowanymi w Rozporządzeniu ministra spraw wewnętrznych i administracji z 17 listopada 2011 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów ogólnogeograficznych, autorzy opracowali bazę pochodną o poziomie uogólnienia odpowiadającego mapom 1:250 000 (określane dalej jako TOPO250). Zaletą takiego rozwiązania (zamiast np. wykorzystania istniejących dla wybranych obszarów baz BDO) jest też zachowanie spójności modelu pojęciowego i klas obiektów między poziomami szczegółowości, co zapewnia wielorodzicielski charakter danych i wieloreprezentację występujących obiektów (rys. 7). Ścisłe określona metodyka generalizacji, pozwala też na określenie relacji między klasami obiektów (i obiektami) na różnych poziomach szczegółowości.

Dzięki temu, możliwe było podjęcie próby zakrojonej na szerszą skalę analizy wpływu sposobu generalizacji informacji geograficznej na poziom szczegółowości, adekwatność i homogeniczność uzyskiwanych wyników. Celem prowadzonych badań było uzyskanie odpowiedzi na pytanie: czy analizując dane o różnej szczegółowości uzyskamy analogicznie zróżnicowane pod względem poziomu uogólnienia pojęciowego, czy wyniki będą spójne niezależnie od poziomu szczegółowości danych źródłowych, czy też otrzymane efekty i rezultaty będą chaotyczne.

### Dane



Rys. 5. Rozmieszczenie obszarów testowych

Testy przeprowadzono na terenach zróżnicowanych pod względem charakteru terenu (stopnia urbanizacji, rzeźby terenu itp.) – rozmieszczenie obszarów testowych przedstawia rysunek 5.

Widoczna spójność między TOPO10 i TOPO250 (rys. 6) oraz znaczące różnice pomiędzy TOPO250 a BDO, mimo tego samego poziomu szczegółowości (różnice w reprezentacji obszaru zabudowanego, obszarów zalesionych, hydrografii itp.) sprawia, iż analiza danych topograficznych TOPO10 może być odniesiona do TOPO250, jest zaś niemal niemożliwe ekstrapolowanie uzyskanych wyników do modelu BDO.

### **Przeprowadzone analizy**

Wykonane analizy polegały na selekcji określonych obiektów i kwantyfikacji ich parametrów, takich jak: liczebność, powierzchnia czy długość. Zdefiniowano zarówno pytania w języku naturalnym, jak i ich odpowiedniki wyrażone w języku SQL. Przeprowadzono selekcję ze względu na klasy obiektów, ich atrybuty, jak i położenie (z wykorzystaniem operatorów przestrzennych). Te same pytania postawiono na różnych poziomach uogólnienia danych (TOPO10 i TOPO250) i porównano uzyskane odpowiedzi.

### **Realizacja eksperymentu**

Przykładowe zapytania do baz danych, zarówno w języku naturalnym jak i sformalizowane za pomocą języka SQL, a także wyniki tych analiz dla obu baz i ich liczbowe porównanie przedstawiono w tabeli 1.

#### **Analiza obiektów liniowych**

Analiza uzyskanych wyników stanowi najistotniejszy element procesu poznawczego i pozwala, bez straty ogólności rozumowania, na rozszerzenie interpretacji jednostkowych współczynników na poszczególne klasy obiektów i bazy danych. Przykładowo pytanie o długość szarp i wałów przeciwpowodziowych (pytanie 1 w tab. 1, górny przykład na rys. 8) przynosi, w zależności od analizowanego poziomu uogólnienia, bardzo różne wyniki. Różnica sięgająca 60% wartości długości wałów dla TOPO10 i TOPO250 wynika z przeprowadzonego procesu generalizacji, który obejmował zarówno odrzucenie zbyt krótkich obiektów, jak i uproszczenie kształtu pozostających obiektów. W porównaniu z pytaniem o długość innego obiektu liniowego – rzeki Jeziorki (pytanie 3 w tab. 1, dolny przykład na rys. 8) gdzie otrzymane wartości różnią się zaledwie o ok. 5% wartości początkowej (z TOPO10), różnica jest tu znacząca. W drugim przypadku mamy do czynienia wyłącznie z uproszczeniem kształtu, cała rzeka Jeziorka jest nadal reprezentowana w bazie TOPO250. W sposób jeszcze bardziej jaskrawy wpływ selekcji na wynik analizy widoczny jest w przypadku linii kolejowych (pytanie 5 w tab.1). Pozostawienie wyłącznie kluczowych linii w bazie TOPO250 powoduje wirtualne „zmniejszenie ich długości” o ponad 300 km, czyli blisko 85% tych modelowanych w TOPO10.

Powyższe przykłady pokazują, jak bardzo proces generalizacji i modelowanie na różnych poziomach szczegółowości, wpływa na wynik zapytań. Nawet pobieżna analiza wskazuje, że to proces selekcji jest kluczowy dla zmiany wyniku analiz długości obiektów liniowych na różnych poziomach uogólnienia. Upraszczanie kształtu nie wpływa na otrzymywane wyniki aż tak istotnie, ponadto wpływ ten jest zależny od charakteru obiektu, który jest upraszczany (stopnia złożoności jego kształtu na źródłowym poziomie szczegółowości). Ten drugi czynnik może być także, choćby częściowo uwzględniany w analizie przestrzennej, np. poprzez wyznaczenie i zastosowanie wymiaru fraktalnego poszczególnych obiektów (Olszewski, 2001).

#### **Przykłady analiz obiektów powierzchniowych**

Mniej zróżnicowane wyniki daje analiza powierzchni obiektów zapisanych w bazie danych, jako poligony. Przykład 2 w tabeli 1 oraz środkowy przykład rysunku 8 pokazują, że

**Tabela 1.** Zestawienie przykładowych zapytań i wyników analiz dla różnych poziomów skalowych

Lp.	Arkusze	Pytanie	SQL	TOPO10	TOPO250	Różnica	
1	Brzeg Dolny	Wybierz wszystkie skarpy/wały przeciwpowodziowe. Jaka jest łączna długość tych obiektów?	select sum(dl) from BBZM_L_BrzegDolny into BrzegDolny_sql01	61,79 km	20,7 km	<b>41 km</b>	<b>64%</b>
2	Brzeg Dolny	Jaka jest całkowita powierzchnia obszarów leśnych na analizowanym obszarze?	select sum(pow) from PKLA_A_BrzegDolny into BrzegDolny_sql02	3305,14 ha	3428 ha	123 ha	3,7%
3	Brzeg Dolny	Jaka jest całkowita długość rzeki Jeziorka?	select sum(dl) from SWRK_L_BrzegDolny where ID_CIEKU="2331" into BrzegDolny_sql03	18,72 km	17,7 km	1 km	5,3%
4a'	Gdynia	Wybierz tereny przemysłowe. Jaka jest powierzchnia tych terenów?	select sum(pow) from PKZB_A where RODZAJ_ZAB="4" into Gdynia_sql01a	350,01 ha	192 ha	<b>158 ha</b>	<b>45,1%</b>
4a"	Gdynia	Wybierz tereny przemysłowe. Jaka jest powierzchnia tych terenów?	select sum(pow) from PKZB_A where RODZAJ_ZAB="4" into Gdynia_sql01a	350,01 ha	0 ha	brak reprezentacji	100%
4b	Gdynia	Ile budynków się na nim znajduje?	Select count(*) from BBBD_A, Gdynia_sql01a where BBBD_A.Obj Within Gdynia_sql01a.Obj into Gdynia_sql01b	1355	0	brak reprezentacji	100%
5	Gdynia	Jaka jest łączna długość linii kolejowych?	SQL TBD: select sum(pow) from SKKL_L	379,72 km	58,4 km	<b>321,3 km</b>	<b>84,6%</b>
6	Warszawa	Liczba budynków gospodarczych?	select count(*) from BBBD_A_Warszawa where FUNKCJA_OG="g" into Warszawa_sql01	25779	0	brak reprezentacji	100%
7	Warszawa	Długość krajowych odcinków jezdni w km?	select sum(dl) from SKJZ_L_Warszawa where KAT_ZARZ = "K" into Warszawa_sql02	120,83 km	0 km	brak reprezentacji	100 %

Oznaczenia przykładów opisanych w kolejnych podrozdziałach: kolor jasnoszary – Analiza obiektów liniowych, ciemnoszary – Przykłady analiz obiektów powierzchniowych, biały – Obiekty niereprezentowane na wyższym poziomie uogólnienia.



różnica w powierzchni może być niewielka, mimo znacznego uproszczenia kształtów (tu obszarów leśnych) oraz selekcji wyłącznie poligonów o powierzchni większej niż przyjęte minimum. Z kolei selekcja obszarów zabudowy przemysłowej (pytanie 4a' w tab. 1) pokazuje sytuację odmienną, kiedy różnica powierzchni w stosunku do „oryginalnej” (z TOPO10) sięga 45%, co wynika z rozdrobnienia wybieranych obszarów, które zanikają w wyniku uogólniania informacji geograficznej. Przy czym warto zauważyć, że powyższe stwierdzenie jest prawdziwe przy założeniu zachowania podziału charakteru obszarów zabudowanych na poziomie TOPO250, scenariusz alternatywny (pytanie 4a'' w tab. 1) omówiono w następnym podrozdziale.

Warto zauważyć, że różnice powierzchni uzyskiwane jako rezultat analiz wynikają w dużej mierze z przyjętego sposobu upraszczania poligonów. W przeciwieństwie do upraszczania linii, która zawsze powoduje jej skrócenie, generalizacja poligonów może powodować zarówno zwiększenie, jak i zmniejszenie powierzchni obiektu (rys. 4). Istotne znaczenie może mieć tu więc wybrany algorytm upraszczania. Wyników bliższych oryginalnym należy spodziewać się na skutek działania algorytmów, które w ramach ewaluacji w trakcie ich opracowywania, jako jeden z czynników uwzględniały zachowanie powierzchni upraszczanego poligonu. W przypadku klasy takiej jak pokrycie terenu (PK w ramach struktury TBD/BDOT10k) dodatkowo na wynik wpływać mogą też priorytety klas w momencie nakładania na siebie poszczególnych, już uproszczonych, rodzajów pokrycia terenu (Fiedukowicz, Olaszewski, 2011). Znaczenie kolejności nakładanych warstw może być nawet istotniejsze niż zastosowana metodyka upraszczania poligonów.

Istotne jest także pytanie, czy znając algorytmy wykorzystywane w generalizacji między poszczególnymi poziomami skalowymi, jesteśmy w stanie przewidzieć zmianę wyniku analizy. Powyższe przykłady pokazują duży rozrzut w wartościach uzyskiwanych różnic – sięgają one od kilku, do kilkudziesięciu procent. Otwarte pozostaje pytanie na ile te zmiany wynikają w istocie ze zmiany poziomu modelowania, na ile zaś są artefaktem, tj. efektem działania algorytmów generalizacyjnych na obiektach. Czy jesteśmy w stanie przewidzieć zmiany wyników analiz między poziomami szczegółowości? Byłoby to cenne ze względu na możliwość wieloskalowego wnioskowania o przestrzeni geograficznej na podstawie analizy monoskalowych danych źródłowych. Jednym z rozwiązań wydaje się ewaluacja procedur generalizacyjnych poprzez ocenę wyników analiz prowadzonych na obiektach źródłowych i zgeneralizowanych. Pozwoliłoby to zachować (do pewnego stopnia) ilościowe charakterystyki takie jak np. procentowy rozkład pokrycia terenu na danym obszarze. Jednocześnie zaś, dzięki statystycznej ocenie algorytmu na etapie jego ewaluacji, pozwoliłoby przewidzieć, że na przykład w wyniku przejścia z poziomu szczegółowości TOPO10 do TOPO250 powierzchnia lasów nie zmieni się o więcej niż 10% z prawdopodobieństwem 95%.

### **Obiekty niereprezentowane na wyższym poziomie uogólnienia**

W dwóch powyższych podrozdziałach, w sposób świadomy, pominięto przykłady związane z obiektami, które nie występują na wyższym stopniu uogólnienia, bądź też zmieniają swoją reprezentację. W takich przypadkach wynik tradycyjnych analiz (takich jak przeprowadzone w ramach eksperymentu) jest jednoznaczny: w jednej z baz przyjmuje konkretną wartość, w drugiej zaś wynosi „0” z uwagi na brak danej klasy obiektów w modelu na innym stopniu uogólnienia. Przy czym o ile we wskazanych tu przykładach (pytanie 4a'', 4b, 6, 7 tab. 1) wynik „0” otrzymujemy dla wyższego stopnia uogólnienia, w ogólnym przypadku

może się też zdarzyć odwrotnie (np. pytanie o długość dróg przyniesie odpowiedź w bazie TOPO250, ale da „0” dla bazy TOPO10, w której modelowane są odcinki jezdni, nie zaś drogi). Różnice w odpowiedziach zależą bowiem od sposobu modelowania w obu bazach.

O ile można dyskutować na temat tego, jak interpretować różnice w wynikach analiz, wynikające z generalizacji obiektów w danej klasie (podrodzinały dotyczące obiektów liniowych i obiektów powierzchniowych), o tyle „znikanie” całych klas obiektów i „zerowe” odpowiedzi mogą budzić pewne kontrowersje. Pokazuje to, że narzędzia analiz zaprojektowane do standardowych baz danych, a zastosowane do dzisiejszych, wielorodzielczych, nie do końca spełniają swoje zadanie i wymagają modyfikacji.

Modyfikację tego typu mógłby stanowić np. mechanizm, który wskazuje twórcy zapytania klasy obiektów powiązane z wybraną klasą (odpowiadającą jej) na innych poziomach szczegółowości. Takie narzędzie byłoby stosunkowo łatwe w implementacji, przy założeniu rzeczywistej wieloreprezentacyjności bazy danych, a więc przechowywaniu w niej także informacji o zależnościach między obiektami (i klasami obiektów) na różnych poziomach szczegółowości. Narzędzie takie mogłoby podpowiadać użytkownikowi alternatywne scenariusze obliczeń z zachowaniem zasady, że interesują nas obiekty w bazie reprezentujące ten sam rzeczywisty obiekt.

Innym rozwiązaniem jest zapisywanie cech geometrycznych i atrybutowych poszczególnych obiektów, zgromadzonych w bazie TOPO10 jako swego rodzaju metadanych, określających charakterystykę obiektów na wyższym poziomie uogólnienia pojęciowego. Rozszerzenie mechanizmów selekcji SQL o wykorzystywanie w trakcie kwerendy realizowanej na poziomie TOPO250 metadanych wyznaczonych w bazie TOPO10, pozwoliłoby na udzielenie dualnej odpowiedzi: wynikającej z analizy geometrii uproszczonej i zintegrowanej oraz pierwotnej geometrii obiektów źródłowych. Pozwoliłoby to użytkownikom dysponującym wyłącznie bazą danych, zawierającą uogólnioną informację geograficzną, na bardziej świadome wnioskowanie, a co za tym idzie na pełniejsze poznanie przestrzeni geograficznej.

## Podsumowanie i wnioski

Wywodząca się z kartograficznej metody badań, metodyka prowadzenia złożonych analiz przestrzennych, została zaimplementowana w wielu narzędziach geoinformacyjnych i stanowi istotny wkład kartografii w rozwój technik badawczych, stosowanych powszechnie w systemach będących częścią składową IIP. Analiza danych przestrzennych, tworzenie modeli prognostycznych i ich walidacja są zaś jednym z podstawowych celów wdrażania infrastruktury geoinformacyjnej.

Zdaniem autorów należy uznać, iż dobrze określona metodyka prowadzenia analiz przestrzennych w oparciu o język SQL rozszerzony o zastosowanie operatorów przestrzennych, dobrze spełnia swą rolę dla monoskalowej bazy danych geograficznych. Rozwój technologii cyfrowej i metodyki modelowania kartograficznego w wieloreprezentacyjnej bazie danych, pozwala jednak na poszerzenie możliwości analitycznych informacji geograficznej. Zastosowanie informacji „źródłowej” o cechach geometrycznych i atrybutowych obiektów, zapisanych w bazie TOPO10 i wykorzystanie ich jako metadanych dla analizy realizowanej w bazie TOPO250, pozwoliłoby na uzyskanie pełniejszej wiedzy o przestrzeni geograficznej. Wiedzy, odnoszącej się nie do *długości kresek na mapie*, lecz do relacji pomiędzy rzeczywistymi obiektami topograficznymi i służącej poznaniu modelowanej w bazie fizycznej przestrzeni. Dane przestrzenne są warte tyle, ile abstrahowana na ich podstawie informacja (Fieduko-

wicz, Kołodziej, Kowalski, Olszewski, 2013), umożliwiającą akwizycję wiedzy i budowę infrastruktury geoinformacyjnej (Olszewski, 2013). Dane topograficzne są więc nie tylko źródłem referencji geometrycznej dla systemów branżowych, lecz przede wszystkim źródłem wiedzy o otaczającym świecie. To wciąż najważniejsza rola jaką pełni topografia, zarówno ta „klasyczna”, jak i ta „współczesna”.

### Literatura

- Bac-Bronowicz J., Olszewski R., 2008: Wielorozdzielcza baza danych jako podstawa georeferencyjna opracowań regionalnych w Polsce, Ogólnopolska Konferencja Kartograficzna.
- Bac-Bronowicz J., Olszewski R., 2009: Możliwości wymiany informacji geograficznej pomiędzy wybranymi zasobami administracji publicznej. *Roczniki Geomatyki*, t. 7, PTIP Warszawa.
- Berlant A.M., 1973: Mapy wzajemnego powiązania zjawisk i ich zastosowanie w badaniach geograficznych [W:] Kartograficzna metoda badań w geografii, IG PAN, z. 3/4, Warszawa.
- Berlant A.M., 1973: Problemy teorii wykorzystania map w badaniach naukowych. [W:] Kartograficzna metoda badań w geografii, IG PAN, z. 3/4, Warszawa.
- Berlant A.M., 1978: Kartograficzeskij metod issledowanija, Izd. Moskowskogo Uniwersiteta, Moskwa.
- Czerny A., 1990: Orientacje problemowe w kartografii teoretycznej, *Polski Przegląd Kartograficzny*, t. 22, nr 4.
- Date C.J., Darwen H., 1997: A Guide to the SQL standard : a user's guide to the standard database language SQL, 4th ed., Addison Wesley.
- Fiedukowicz A., Kołodziej A., Kowalski P., Olszewski, 2013: Społeczeństwo geoinformacyjne i przetwarzanie danych przestrzennych. [W:] Olszewski R., Gotlib D. (red.) Rola bazy danych obiektów topograficznych w tworzeniu infrastruktury informacji przestrzennej w Polsce, GUGiK.
- Fiedukowicz A., Olszewski R., 2011: Ewaluacja statystyczna jako miara poprawności generalizacji informacji geograficznej na przykładzie opracowania komponentów pochodnych BDG. [W:] Zastosowanie statystyki w GIS i kartografii, Główne problemy współczesnej kartografii: 104-126.
- Gotlib D., Iwaniak A., Olszewski R., 2007: GIS. Obszary zastosowań, Wydawnictwo Naukowe PWN, Warszawa.
- Grygorenko W., 1984: Ewolucja poglądów na temat roli i funkcji kartografii, *Polski Przegląd Kartograficzny*, t. 16, nr 2.
- Makowski A., 2005: System informacji topograficznej kraju. Teoretyczne i metodyczne opracowanie koncepcyjne. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej.
- Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W., 2006: GIS, Teoria i praktyka, Wydawnictwo Naukowe PWN, Warszawa.
- Olszewski R., 2001: Zastosowanie analizy fraktalnej w kartograficznej metodzie badań wybranych komponentów środowiska przyrodniczego. Rozprawa doktorska, opiekun prof. dr hab. inż. Andrzej Makowski, Wydział Geodezji i Kartografii Politechniki Warszawskiej.
- Olszewski R., 2009: Bazy danych przestrzennych jako element infrastruktury geoinformacyjnej w Polsce. [W:] Żyszkowska W., Spallek W., Bazy danych w kartografii, Uniwersytet Wrocławski, Wrocław.
- Olszewski R., 2013: Wpływ dyrektywy INSPIRE na rozwój kartografii w Polsce. [W:] Olszewski R., Gotlib D. (red.) Rola bazy danych obiektów topograficznych w tworzeniu infrastruktury informacji przestrzennej w Polsce, GUGiK.
- Saliszczew K.A., 1955: O kartograficzeskome metodie issledowanija, *Wiestnik Moskowsogo Uniwersiteta*, Sieria fiziko-mat., z 10.
- Saliszczew K.A., 1968: Mietodika analiza pri issledowanii jawlenij po kartam. *Wiestnik Moskowsogo Uniwersiteta*, Sieria V, *Gieografija*, vol. 6.
- Saliszczew K.A., 1998: Kartografia ogólna. Wydawnictwo naukowe PWN, Warszawa.
- Weibel R., 1991: Amplified intelligence and rule-base systems. [In:] Buttenfield B., McMaster R., (red.) *Map generalization: making rules for knowledge representation*, Longman, London.
- Weibel R., 1995: Map generalization in the context of digital systems, *Cartography and GIS* vol. 22, no. 4.
- Rozporządzenie MSWiA, 2011b: Rozporządzenie ministra spraw wewnętrznych i administracji z 17 listopada 2011 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów ogólnogeograficznych, a także standardowych opracowań kartograficznych. Dz.U. 2011, nr 279, poz. 1642.

**Abstract**

*For ages maps have been the tool for modeling and representation of geographic space. As cartographic techniques were developed, maps became useful in much wider aspects than only for the original function of object localization. The Cartographic Research Method allows studies on physical space using map, treated as model of this space.*

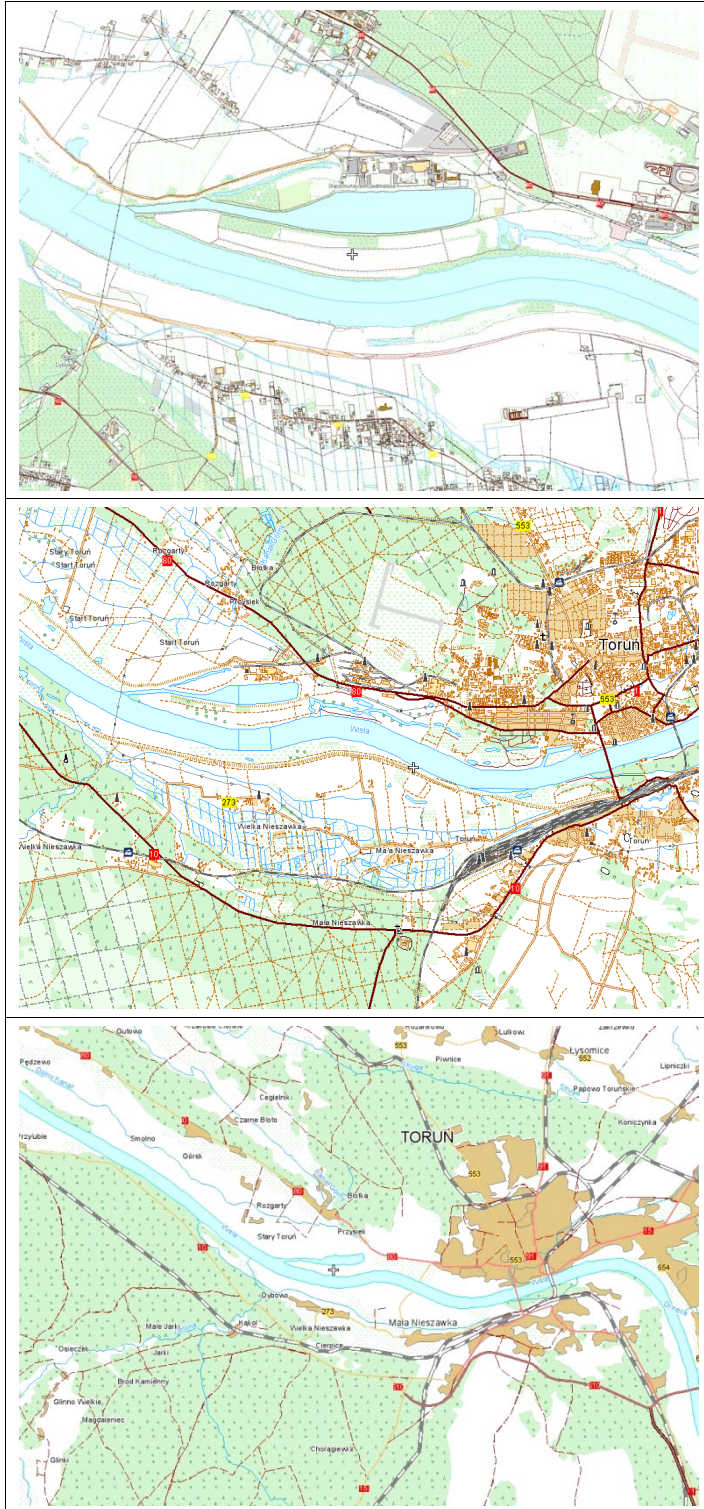
*One of the basic information tools used in GIS technology is SQL (Structured Query Language), language which is used for creation of formalised queries allowing for selection, aggregation and analysis of information collected in database. Application of SQL is also used to learn about characteristics of selected objects which can come from both geometry and attributes ascribed to them. Therefore, the use of spatial data analysis by applying SQL allows for indirect studies on geographic space.*

*Results of analysis of geographic space are closely connected with the level of detail (LoD) or scale of the source data. The authors' goal was to check how reliable are results received by analyzing different LoD of multiresolution topographic database.*

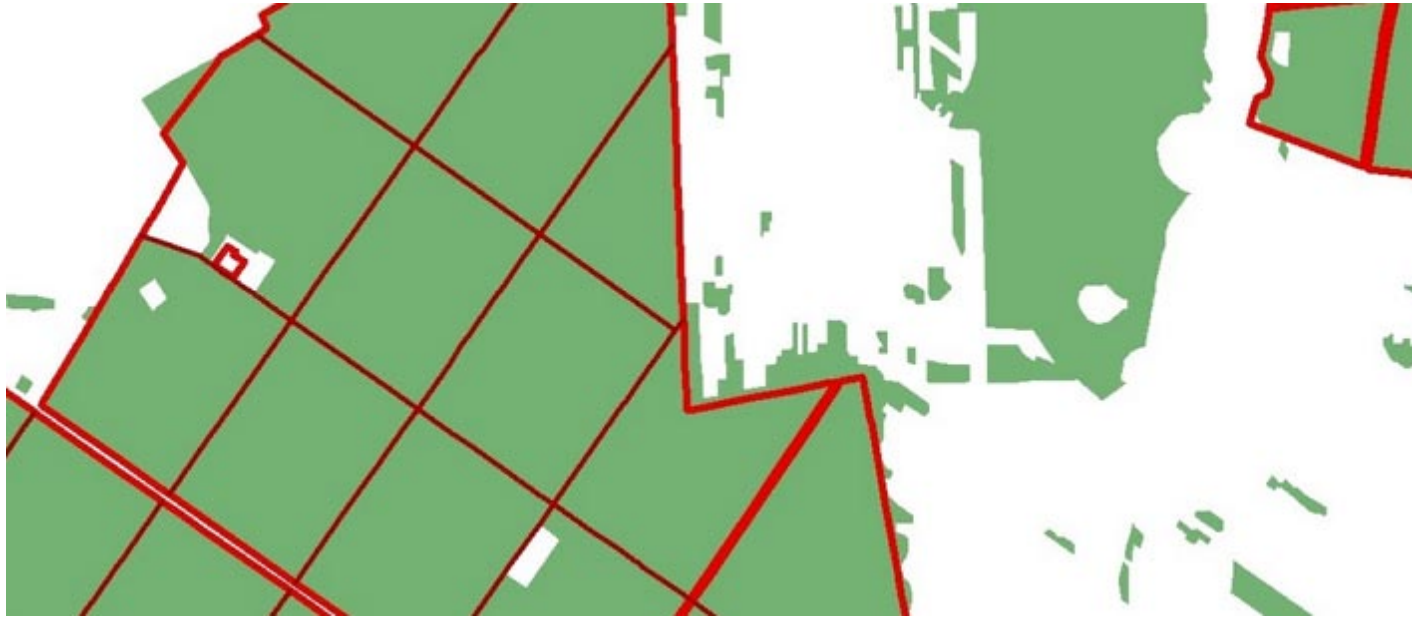
*The authors used the "classic methodology" of Cartographic Research Method by applying modern analytical tools in GIS. Unified analysis were used for various LoD (TOPO10 and TOPO250) and different test areas. Obtained results show that changing LoD may strongly influence the results of the analysis, which can lead to their wrong evaluation and interpretation.*

dr hab. inż. Robert Olszewski, prof. PW  
R.Olszewski@gik.pw.edu.pl

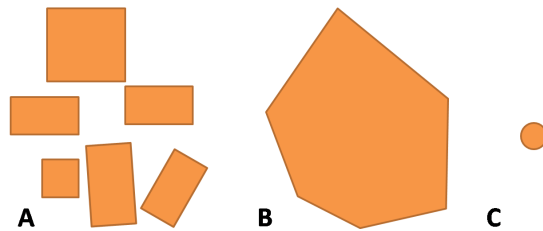
mgr inż. Anna Fiedukowicz  
A.Fiedukowicz@gik.pw.edu.pl



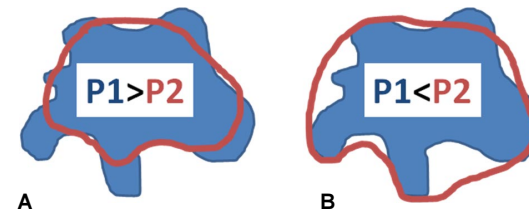
**Rys. 1.** Porównanie stopnia  
szczegółowości  
danych TBD, VMap L2 i BDO  
(źródło: geoportail.gov.pl)



**Rys. 2.** Porównanie zasięgu obszarów leśnych w ramach BDOT10k (kolor zielony) oraz Leśnej Mapy Numerycznej – LMN (czerwone granice). Różnice wynikają z różnego zdefiniowania lasu w obu bazach: w LMN zastosowano definicję prawną wynikającą z ustawy o lasach, w BDOT10k wyznaczanie obszarów leśnych oparte jest o fizjografię terenu

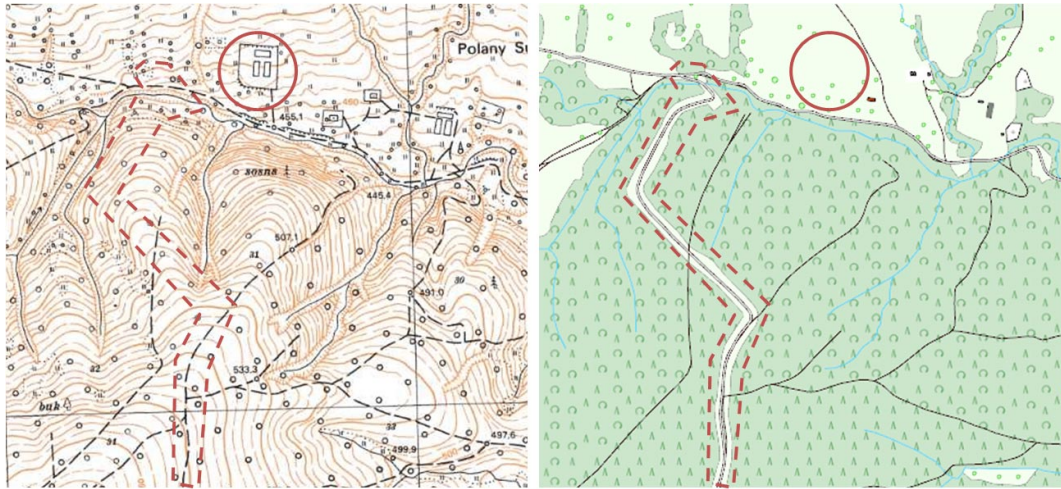


**Rys. 3.** Reprezentacja przestrzeni miejskiej na różnych poziomach szczegółowości: A – zbiór budynków, B – obszar zabudowany, C – reprezentacja punktowa



**Rys. 4.** Upraszczenie poligonu w wyniku generalizacji: A – zmniejszanie powierzchni poligonu, B – zwiększanie powierzchni poligonu

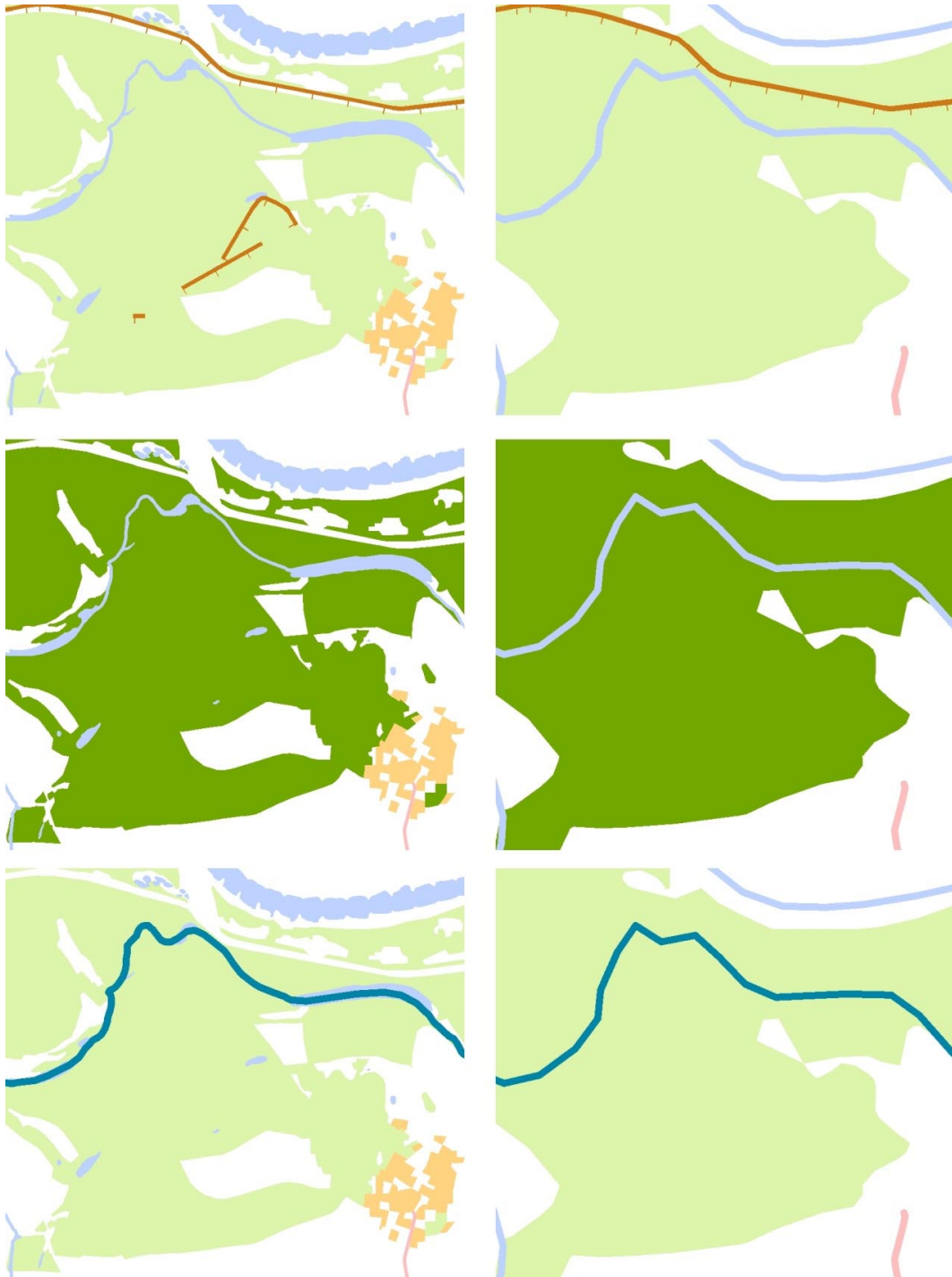




**Rys. 6.** Różnice w aktualności dawnej mapy topograficznej (1:10 000) i obecnej bazy BDOT10k. Zmiany w czerwonych obrysach: linią ciągłą zaznaczono nieistniejące już budynki, przerywaną zaś nowopowstałą drogę, wraz z wyciętym pasem lasu (źródło: geoportal.gov.pl)



**Rys. 7.** Od lewej: wizualizacja wybranych klas obiektów bazy TOPO10, TOPO250 (wynik generalizacji TOPO10) oraz BDO



**Rys. 8.** Wizualne porównanie wyników analiz dla obszaru testowego Brzeg Dolny: po lewej 1: 10 000, po prawej 1: 250 000; od góry w kolejnych rzędach analizy 1, 2, 3 z tabeli 1