

Rozdział 2

Dane przestrzenne

Niezależnie od tego w jaki sposób wykorzystujemy GIS dane są jego niezbędnym elementem. Dziedziną i związanymi z nią zawodami, która jest postrzegana jako symbol współczesności jest obecnie **data science**. Pozwala ona nie tylko pozyskiwać z danych informację ale wykorzystując techniki sztucznej inteligencji także wiedzę, czyli wykrywać jakie są mechanizmy zachodzenia różnorodnych złożonych procesów. Dziedzina ta wykorzystuje ogromne ilości danych, które cały czas są tworzone za pomocą najróżnorodniejszych technik. Dane te aby mogły być przetwarzane przez uniwersalne narzędzia i metody muszą być przedstawiane w uniwersalny sposób. Jednym z powszechnie stosowanych rozwiązań jest najprostszy z możliwych sposobów – tekstowe pliki formatu **CSV** (comma-separated values – wartości rozdzielone przecinkiem).

```
kraj,piwo_p,wodka_p,wino_p,litry_alkohol,kontynent
Afghanistan,0,0,0,0.0,Asia
Albania,89,132,54,4.9,Europe
Algeria,25,0,14,0.7,Africa
Andorra,245,138,312,12.4,Europe
Angola,217,57,45,5.9,Africa
Antigua & Barbuda,102,128,45,4.9,North America
Argentina,193,25,221,8.3,South America
Armenia,21,179,11,3.8,Europe
Australia,261,72,212,10.4,Oceania
Austria,279,75,191,9.7,Europe
```

Rysunek 2.1 Przykładowy plik tekstowy z danymi formatu **CSV**. Dane opisują różne rodzaje spożycia alkoholu w poszczególnych krajach. Każdy wiersz zawiera kolejno następujące dane: nazwę kraju, roczną liczbę porcji piwa na mieszkańca, porcji wódki, porcji wina, roczne spożycie litrów czystego alkoholu przez mieszkańca, nazwę kontynentu. Dodatkowo pierwsza linia zawiera nazwy poszczególnych zmiennych.

Każdy wiersz opisuje oddzielną obserwację z kompletem pozyskanych pomiarów (zmiennych, atrybutów). Te atrybuty każdej obserwacji mogą mieć charakter (typ) tekstu, liczby całkowitej lub dziesiętnej (zamiast przecinka używamy

kropki). Tak przygotowane dane mogą być w prosty sposób stabilizowane co jest zwykle pierwszym krokiem polegającym na wprowadzeniu danych do środowiska analitycznego. Możemy je wprowadzić także do GIS i przedstawić w postaci tablicy. **Tablica danych tekstowych jest rodzajem danych stosowanych w GIS** (choć dane te nie mają charakteru przestrzennego).

	kraj	piwo_p	wodka_p	wino_p	litry_alkohol	kontynent
1	Colombia	159	76	3	4.2	South America
2	Chad	15	1	1	0.4	Africa
3	Chile	130	124	172	7.6	South America
4	Brunei	31	2	1	0.6	Asia
5	Bulgaria	231	252	94	10.3	Europe
6	Botswana	173	35	35	5.4	Africa
7	Brazil	245	145	16	7.2	South America

Atrybuty mają swój typ danych.

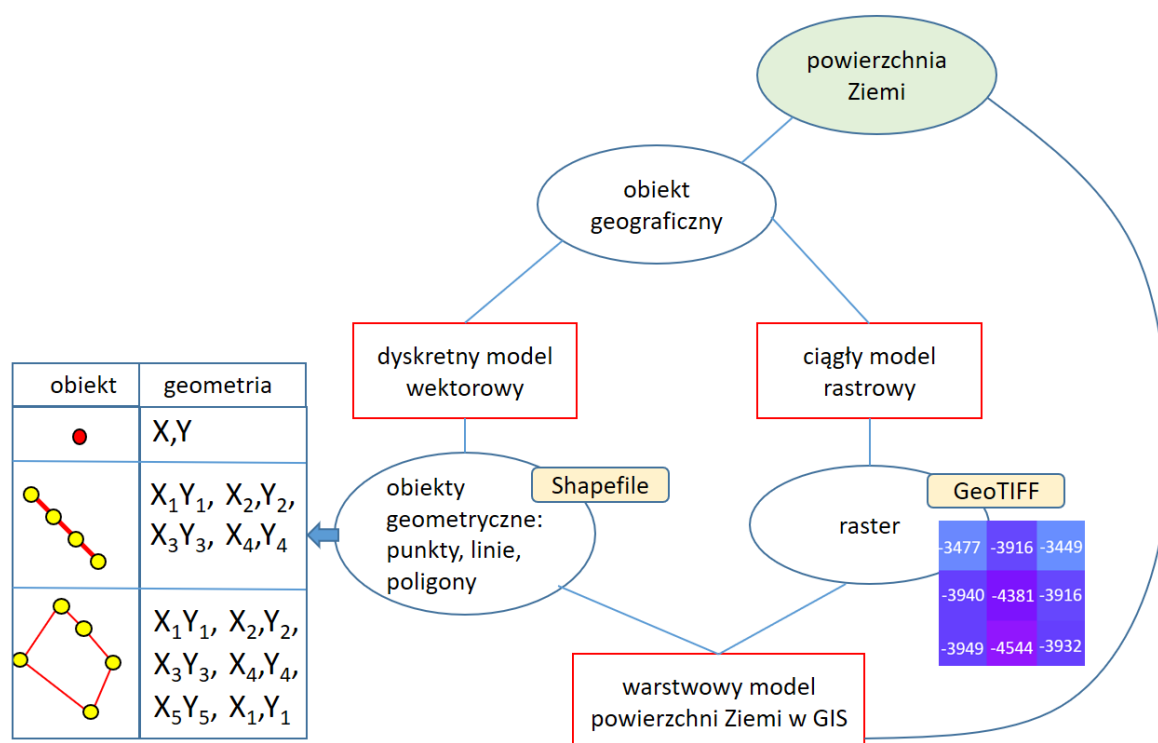
	Id	Name	Alias	Type	Type name	Length	Precision
abc	5	kontynent		QString	text	0	0
1.2	4	litry_alkohol		double	double	0	0
123	3	wino_p		int	integer	0	0
123	2	wodka_p		int	integer	0	0
123	1	piwo_p		int	integer	0	0
abc	0	kraj		QString	text	0	0

Dane organizowane są w tablicach. Każda obserwacja tworzy oddzielny wiersz (**rekord**). Każdy pomiar (atrybut, zmienna) obserwacji jest zapisywany w oddzielnej kolumnie (**polu**).

Rysunek 2.2 Dane tekstowe w GIS w postaci tablicy.

Struktura tych danych polegająca na specyficznej organizacji danych w tablicy jest metodą **data science** zapewniającą uniwersalne przetwarzanie danych. Tak zorganizowane i uporządkowane dane noszą nazwę **tidy data** (tidy można tłumaczyć jako zorganizowane lub uporządkowane). Mają one następujące wspólne cechy. Każda określona zmienna (atrybut) danej obserwacji jest zapisywana w kolumnie. Każda oddzielna obserwacja szeregu zmiennych tworzy oddzielny wiersz. Dla każdego tematu obserwacji tworzona jest oddzielna tablica. Jeżeli posiadamy parę tablic to powinny one zawierać kolumnę, która umożliwi ich łączenie. Nasze obserwacje dotyczą poszczególnych krajów, które mogą być traktowane jako obiekty geograficzne, oczywiście te obiekty w tej tablicy reprezentowane są wyłącznie przez nazwę.

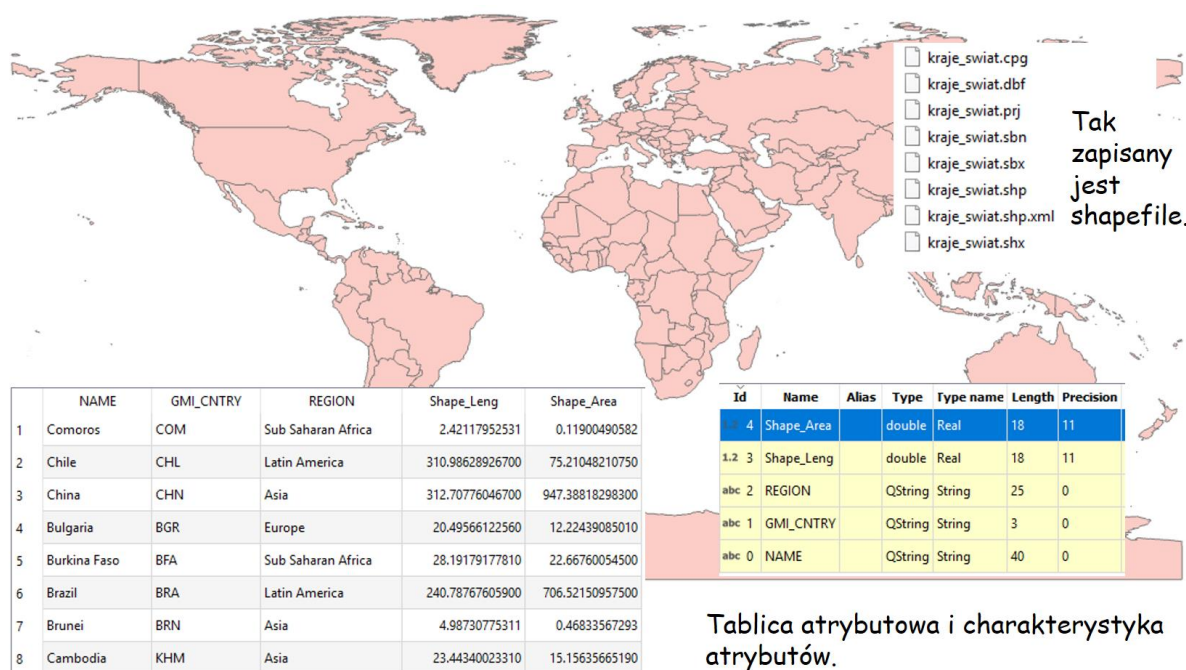
Aby obiekty geograficzne powierzchni Ziemi mogły być przedstawione w postaci danych, potrzebny jest model danych przestrzennych, który pozwoli na ich reprezentację w postaci cyfrowej.



Rysunek 2.3 Model danych przestrzennych stosowany w GIS.

Powierzchnia Ziemi jest pokryta szeroko rozumianymi obiektami geograficznymi. Obiektami mogą być budynki, jeziora, drogi ale także teren i jego ukształtowanie, oraz ocean i jego temperatura powierzchniowa. Obiekty mogą być dzielone na klasy tematyczne (np. rzeki, jeziora, pokrycie terenu, wysokość terenu). Te różne rodzaje obiektów geograficznych można jednak podzielić na dwie różne klasy. Klasę obiektów dyskretnych, które zajmują określone miejsce w przestrzeni i często posiadają wyraźnie określone granice i klasę obiektów ciągłych jak na przykład rozkład opadu deszczu w ciągu roku czy wysokość terenu. Przykładem obiektów dyskretnych będą: studnia, źródło, rzeka, droga, jezioro, obszar powiatu. Ich dobrym reprezentantem będą punkt, linia oraz wielobok (zwany powszechnie w środowisku ludzi zajmujących się GIS polygonem). Geometrie tych obiektów tworzą przestrzenny element wektorowego modelu danych. **Dane wektorowe są pierwszym rodzajem danych przestrzennych wykorzystywanym w GIS.** Geometria punktu na powierzchni Ziemi jest zapisywana jako dwie jego współrzędne (np. długość i szerokość geograficzną). W przypadku konstrukcji geometrycznej linii potrzebujemy szeregu punktów (te konstrukcyjne punkty nazywamy werteksami). Podobnie przy tworzeniu

poligonu (wieloboku) używamy werteksów i aby zamknąć ich ciąg powtarzamy pierwszy werteks. Zostało to przedstawione w tablicy obiektów i ich geometrii na rysunku 2.3. Każdy obiekt geometryczny wraz ze swoimi atrybutami tworzy jeden rekord (wiersz) w tablicy danych (tzw. tablicy atrybutowej). Na rysunku 2.4 przedstawiono zbiór danych przestrzennych tworzących warstwę krajów świata.

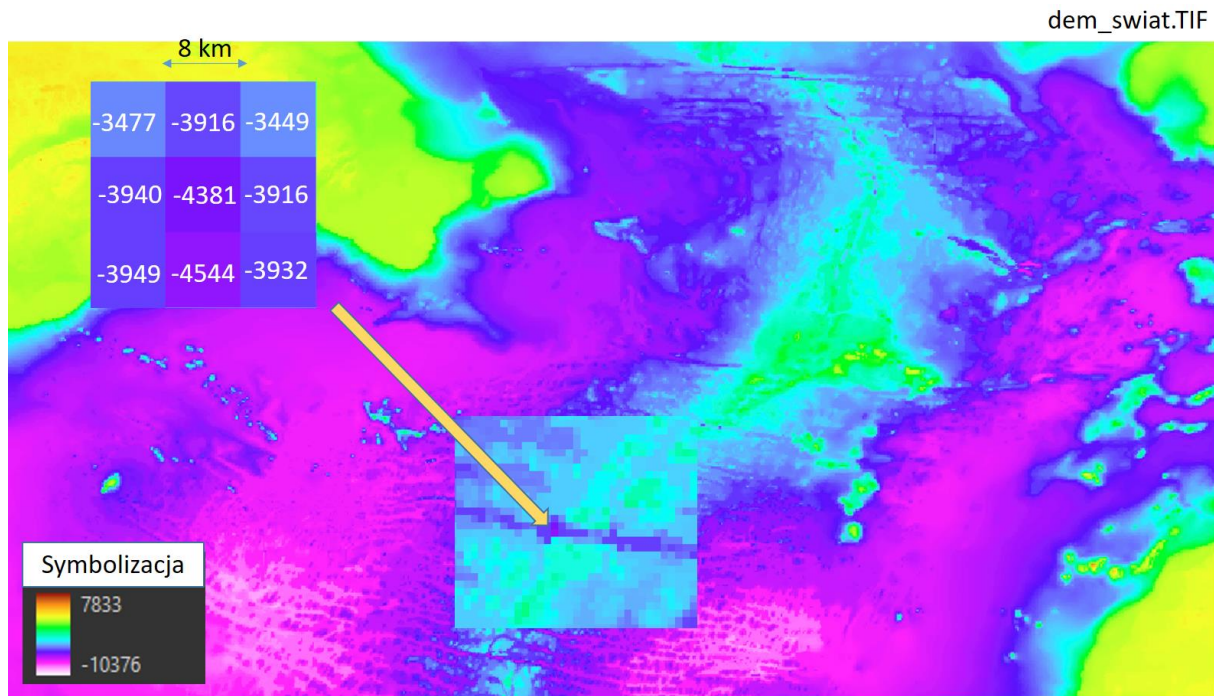


Rysunek 2.4 Wektorowa warstwa poligonów zapisana w formacie Shapefile.

Jest to warstwa poligonów. Warstwy (zbiory danych wektorowych) tworzy się oddzielnie dla punktów, linii i poligonów. Pod względem logicznym każda warstwa składa się z rekordów zawierających geometrie obiektów i związane z nią atrybuty. Często geometria i atrybuty są fizycznie zapisywane oddzielnie. Na rysunku 2.4 przedstawiono najbardziej powszechny format zapisu danych wektorowych **Shapefile**. Warstwa poligonów wraz z atrybutami została zapisana w szeregu plików o wspólnej nazwie i różnych rozszerzeniach. Tablica z atrybutami, która może być traktowana jako samodzielna jest zapisywana z rozszerzeniem **.dbf**. Warstwa posiadająca obiekty geometryczne może być wyrysowana na obszarze roboczym danego programu GIS. Sposób rysowania poszczególnych obiektów nosi nazwę **symbolizacji**. Na rysunku 2.4 wszystkie poligony mają takie samo wypełnienie i obrys (szara linia). Rodzaj symbolizacji może być zależny od wartości w wybranym polu (kolumnie) tablicy atrybutowej.

Obiekty o charakterze ciągłym są reprezentowane za pomocą modelu rastrowego. **Dane rastrowe są drugim podstawowym rodzajem danych przestrzennych używanym w GIS.** Model rastrowy tworzy mozaikę kwadratów (komórek rastra) pokrywających powierzchnię Ziemi. Mozaika tworzy prostokąt

składający się z wierszy i kolumn komórek. Stąd istotną charakterystyką rastra jest **liczba wierszy i kolumn**. Ponieważ komórki pokrywają powierzchnię Ziemi kolejną ważną cechą jest długość ich boku, jest to tak zwana **rozdzielczość przestrzenna rastra**.

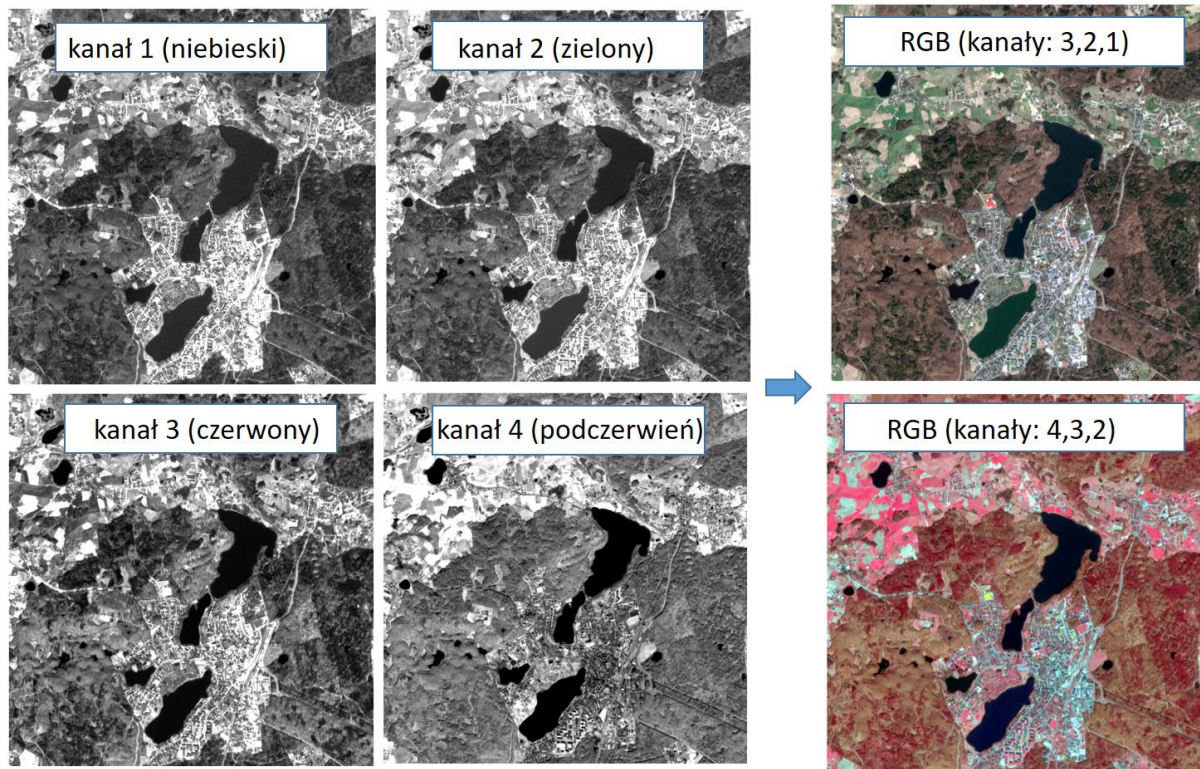


Rysunek 2.5 Rastrowa warstwa głębokości oceanu. Kolejne przybliżenia wycinków rastra pokazują jego strukturę składającą się z kwadratowych komórek.

Każda komórka rastra zawiera jedną wartość. Może to być liczba całkowita tak jak na Rysunku 2.5, gdzie przedstawia ona głębokość w środku komórki, ale może także być identyfikatorem pewnej klasy (np. liczba 0 reprezentuje wodę a, a liczba 1 ląd). Komórki rastrowe mogą także zawierać liczby rzeczywiste. Podobnie jak w przypadku warstw wektorowych wizualizacja rastru wymaga **symbolizacji**, polega ona na przypisaniu określonego koloru lub odcienia szarości do konkretnych wartości lub ich przedziałów na rastrze. Istnieje wiele formatów zapisu danych rastrowych, jednym z najbardziej uniwersalnych jest **GeoTIFF**. W formacie tym zapisuje się rastry jako pojedynczy plik z rozszerzeniem **.tif**, informacja o położeniu w przestrzeni jest zawarta wewnątrz tego pliku.

Ważnym rodzajem danych wykorzystywanymi w GIS są zdjęcia satelitarne. Są one pozyskiwane jako rastry o bardziej złożonej strukturze niż omawiane powyżej. Zdjęcia satelitarne to zestawy dokładnie pokrywających się rastrów tworzonych dla wybranych przedziałów promieniowania (najczęściej odbitego) dochodzącego z powierzchni Ziemi. Rastry te nazywamy kanałami zdjęcia

satelitarne. Może być ich parę, kilkanaście lub nawet kilkaset, ale są często zapisywane jako jeden plik formatu GeoTIFF w postaci tzw. **kompozytu**.



Rysunek 2.6 Cztery kanały zdjęcia satelitarne i symbolizacja ich kompozytu dla dwóch zestawów kanałów.

Na rysunku 2.6 z lewej strony pokazane są cztery kanały zdjęcia satelitarne (są one wyświetlone za pomocą odcieni szarości). Połączono je w kompozyt czterokanałowy. Wizualizacja kompozytu polega na wykorzystaniu modelu kolorów **RGB** (czerwony, zielony, niebieski) do jednoczesnego wyświetlenia trzech kanałów w tych kolorach, w rezultacie tworzy się bogata paleta różnych barw. Jeżeli do RGB przypiszemy odpowiadające im zdjęcia (zrobione w zakresie czerwonym, zielonym, niebieskim) otrzymamy obraz o kolorach zbliżonych do rzeczywistych (tak jak widzi to ludzkie oko). Jeżeli natomiast przypiszemy inny zestaw kanałów (kompozyt dolny) otrzymamy tzw. fałszywe kolory, które mogą nam dostarczyć informacje niewidoczne w innych zestawach kanałów.

Warstwy wektorowe i rastrowe reprezentujące istotne dla danego zagadnienia obiekty geograficzne tworzą **warstwowy model powierzchni Ziemi**.