

Formaty danych – cz. 3 GML

Po zapoznaniu się ze strukturą XML i schematami XSD pora wyjaśnić jaki on ma związek z danymi przestrzennymi. Otóż, jak już było wspomniane, standardem danych przestrzenny w Unii Europejskiej (a więc i w Polsce) jest Geography Markup Language (GML). Jest to zasadniczo XML o ustalonych schematach XSD za pomocą odgórnych specyfikacji. Dzięki niej, systemy informacji geograficznej, na przykład oprogramowanie GIS, jest w stanie wyświetlać dane. Wydziały geodezji jednostek administracyjnych kraju są zobowiązane do przyjmowania, przechowywania i udostępniania na żądanie (za opłatą) danych w formacie GML.

Dlaczego zasadniczo ustala się coś takiego jak standard danych?

Interoperacyjność systemów jest kluczowa dla funkcjonalności systemów, dlatego też systemy muszą być zdolne do komunikacji a format danych nie może ich ograniczać.

Co zawiera GML?

Podczas przenoszenia danych z systemu, czyli zapisywania ich do formatu GML, zawsze zapiszą się nam dwa pliki: *.gml oraz *.xsd. Czyli plik z danymi i plik ze schematem do walidacji.

GML to bardzo rozbudowany format (co za tym idzie, były spore problemy z jego wdrożeniem), który zawierać może kilka typów danych jak:

- obiekty
- geometrię
- układ odniesienia
- topologię
- czas
- pokrycie terenu
- pomiary (obserwacje) wraz z jednostkami
- zasady reprezentacji obiektów (stylizacja)

Weźmy na przykład prosty plik z dwoma punktami:



Proszę otworzyć plik “Punkty.gml”:

Jak widzimy plik zaczyna się od identycznego nagłówka, jak każdy plik XML:

```
<?xml version="1.0" encoding="utf-8" ?>
```

Znacznikiem <root> jest tutaj:

```
<ogr:FeatureCollection  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://ogr.maptools.org/ Punkty.xsd"  
  xmlns:ogr="http://ogr.maptools.org/"  
  xmlns:gml="http://www.opengis.net/gml">
```

Czyli znacznik **FeatureCollection** (zbiór obiektów) o 4 atrybutach, które są odnośnikami do specyfikacji, jak również schematu. Proszę zwrócić uwagę na przestrzeń nazw: ogr to standard zapisu danych wektorowych używany przez bibliotekę GDAL, która jest standardem OGC - Open Geospatial Consortium.

Pod nim znajdują się dane:

- Metadane - układ odniesienia, zakres danych
- Same dane

Na początku pojawiają się metadane:

```
<gml:boundedBy>  
  <gml:Box>
```

```
<gml:coord><gml:X>4344844.915299112</gml:X><gml:Y>75187.80643969306</gml:Y></gml:coord>
```

```
<gml:coord><gml:X>4449374.304739662</gml:X><gml:Y>78244.22133561556</gml:Y></gml:coord>
```

```
</gml:Box>  
</gml:boundedBy>
```

Czyli **BoundingBox** - zakres geograficzny w danym układzie w jakim mieszczą się dane. Jest to najmniejszy prostokąt, który można opisać na danych.

Kolejny znacznik zawiera informacje o samych już danych. Każdy znacznik **“featureMember”** zawiera informacje o kolejnych punktach na warstwie, tu mamy dwa:

```
<gml:featureMember>  
  <ogr:Punkty fid="Punkty.0">
```

```

    <ogr:geometryProperty><gml:Point
srsName="EPSG:2177"><gml:coordinates>4344844.91529911,75187.8064396931</gml:co
ordinates></gml:Point></ogr:geometryProperty>
    <ogr:id>1</ogr:id>
    <ogr:Nazwa>P1</ogr:Nazwa>
    <ogr:Data_pomia xsi:nil="true"/>
    <ogr:Wysokosc>123.320</ogr:Wysokosc>
    </ogr:Punkty>
</gml:featureMember>

```

```

<gml:featureMember>
    <ogr:Punkty fid="Punkty.1">
    <ogr:geometryProperty><gml:Point
srsName="EPSG:2177"><gml:coordinates>4449374.30473966,78244.2213356156</gml:co
ordinates></gml:Point></ogr:geometryProperty>
    <ogr:id>2</ogr:id>
    <ogr:Nazwa>P2</ogr:Nazwa>
    <ogr:Data_pomia xsi:nil="true"/>
    <ogr:Wysokosc>123.333</ogr:Wysokosc>
    </ogr:Punkty>
</gml:featureMember>

```

I znacznik zamykający:

```
</ogr:FeatureCollection>
```

W punktach widzimy takie znaczniki jak:

- **<ogr:Punkty fid="Punkty.0">** - to znacznik posiadający nazwę warstwy wektorowej z atrybutem unikalnie identyfikującym obiekty
- **<ogr:geometryProperty>** To znacznik, który zagnieżdża właściwości geometryczne obiektów, widzimy znaczniki podrzędne:
- **<gml:Point>** - czyli standard już GML, który zawiera informacje o typie geometrii, tutaj Point oznacza punkt. Inne możliwe geometrie to **LineString** (linia), **Polygon** lub **LinearRing** (wielobok) i czasem Coverage (pokrycie terenu) oraz inne, związane z punktem, linią lub wielobokiem, które pojawiają się rzadziej

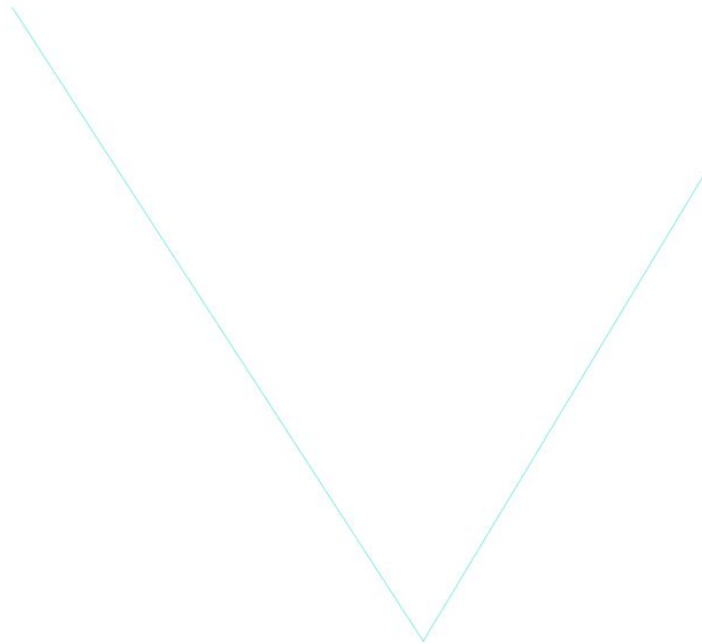
Znacznik point zawiera atrybut **srsName** - określający układ odniesienia. Układ określany jest numerem **EPSG** (European Petroleum Survey Group). To standardowy spis układów odniesienia określanych unikalnym numerem, szybkie wyszukiwanie w Google mówi, że EPSG: 2177 to ETRS89 / Poland CS2000 strefa 6. Numery zwykle są 4-cyfrowe, czasem dłuższe (5 lub 6 cyfr). Domyślny układ (w przypadku braku definicji układu) to EPSG:4326, który odpowiada układowi WGS84 kątowemu.

Dalszy podrzędny znacznik zawiera informację o współrzędnych **<gml:coordinates>** - tu tylko para dwóch współrzędnych (X,Y) i znaczniki zamykające.

Dalsze znaczniki to atrybuty. Ta warstwa posiada atrybuty **id**, **Nazwa**, **Data pomiaru** oraz **Wysokosc**. Znaczniki przechowują te wartości dla każdego obiektu.

Jak widać, znając język XML nietrudno jest zrozumieć sam zapis pliku GML. Każdy plik będzie miał zbliżony wygląd. Oczywiście większe warstwy będą miały odpowiednio większe pliki. Zwróćmy uwagę że plik z dwoma punktami zajmuje 2 kilobajty, mimo, że to niewiele, to rozmiar tego pliku mógłby być znacznie mniejszy. W GML wiele informacji się powtarza (mówimy że są redundantne), co sprawia że pliki te mają bardzo duże rozmiary.

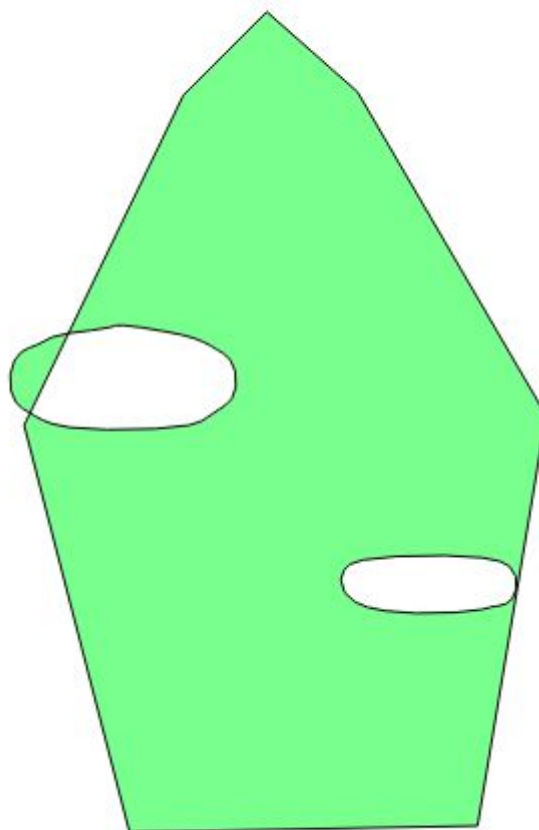
Następny plik jest wielolinia (Plik "linia.gml"):



Plik zawiera identyczne nagłówki jak plik z punktami. Podobnie określony jest prostokąt okalający - oczywiście o innych współrzędnych narożników. Warstwa zawiera tylko jeden obiekt (jeden znacznik featureMember) w którym:

- znajduje się oczywiście **inny identyfikator obiektu**;
- Znajduje się inna geometria typu **<gml:LineString>**, która tym razem składa się z 3 par współrzędnych (proszę zwrócić uwagę, że X i Y są rozdzielone przecinkiem, następujące po sobie współrzędne są rozdzielone spacją, a znakiem dziesiętnym jest kropka);
- Znajduje się jeden atrybut **<ogr:id>22/<ogr:id>** czyli atrybut o nazwie id o wartości 22

Następny plik jest wielobokiem z otworami (Plik "example.gml"):



Proszę otworzyć ten plik.

Poza nagłówkiem, widzimy również znacznik BoundingBox - minimalnego prostokąta okalającego.

Struktura tego pliku jest niemal identyczna do poprzednich plików poza:

- Oczywiście **inną nazwą elementu**
- **Innym układem odniesienia**
- **Inną geometrią**

Geometria tym razem to typ **<gml:Polygon>**. W każdym formacie danych, Polygon jest szczególnie skomplikowanym typem danych, dlatego, że dopuszcza się istnienie wieloboków z "otworami". Tutaj dlatego też pod znacznikiem Polygon, znajduje się podznacznik **<gml:outerBoundaryIs>** - czyli zewnętrzna krawędź wieloboku, która ma geometrie określoną innym typem danej **<gml:LinearRing>** (nie jest to Polygon, ale ma

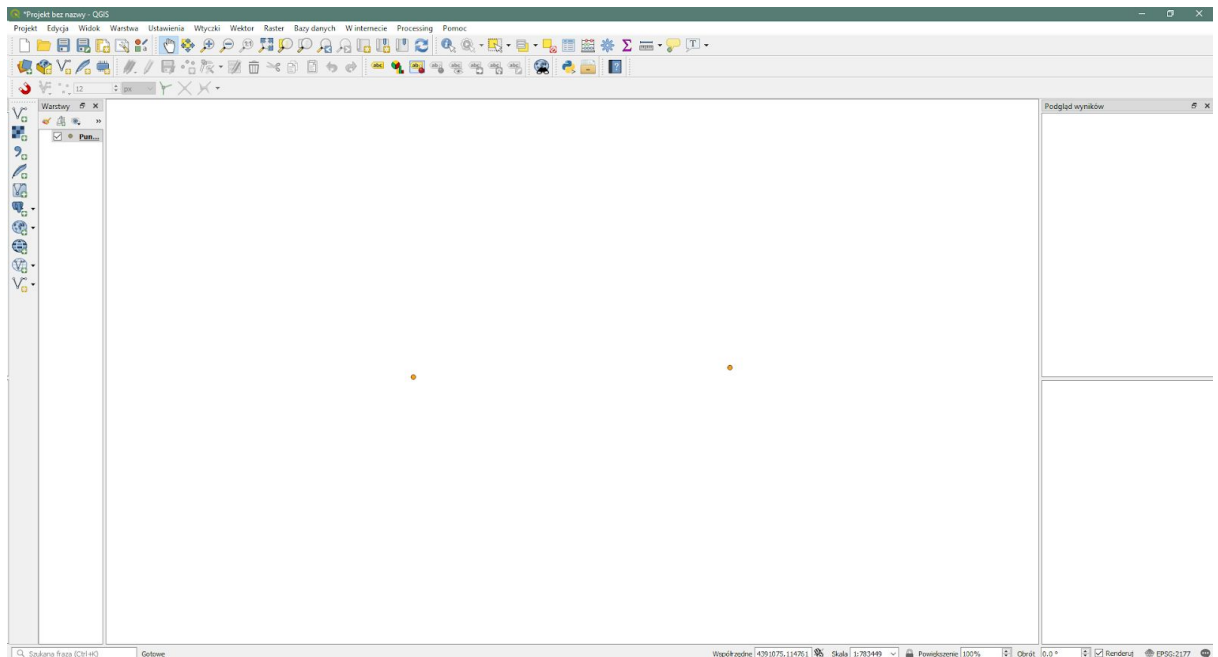
bardzo zbliżone właściwości). Następujące znaczniki **<gml:inneBoundaryIs>** to definicja “otworów” w warstwie (dwóch).

QGIS

Na tych zajęciach korzystać będziemy z oprogramowania GIS o nazwie QGIS. Jest to darmowe oprogramowanie będące systemem informacji przestrzennych, o zbliżonych właściwościach do ArcGIS.

Wykorzystamy go do załadowania i obejrzenia plików GML. Proszę pobrać QGIS w wersji 3.x ze strony: <https://qgis.org/pl/site/>

Po zainstalowaniu na komputerach uruchamiamy oprogramowanie QGIS3:

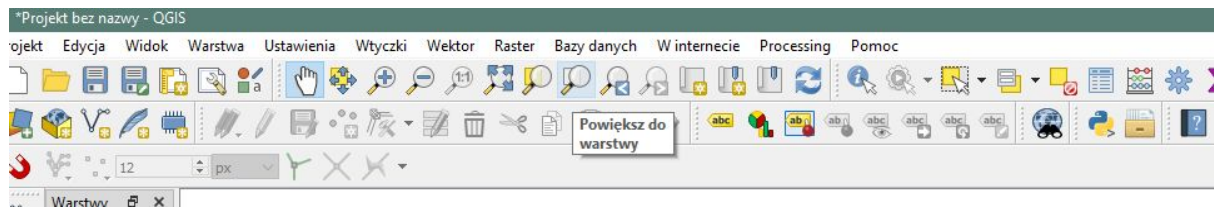


Teraz wystarczy “przeciągnąć” jedną z warstw do okna programu. Weźmy na przykład “Punkty.gml”.

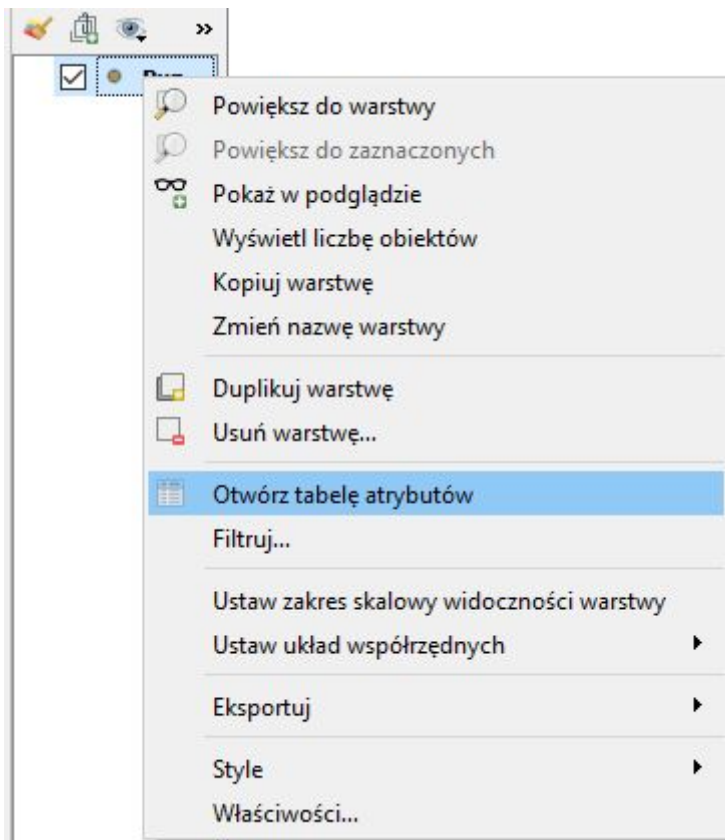
Po lewej stronie w oknie warstwy pojawi się jej nazwa:



W głównym oknie powinny wyświetlić się dwa punkty. Jeśli się nie wyświetliły, klikamy w oknie “Warstwy” raz na nazwę warstwy (nie na zaznaczenie!, to wyłącza wyświetlanie warstwy) i klikamy “powiększ do warstwy”:



Następnie otworzymy tabelę atrybutów warstwy poprzez kliknięciem prawym przyciskiem myszy na nazwę warstwy:

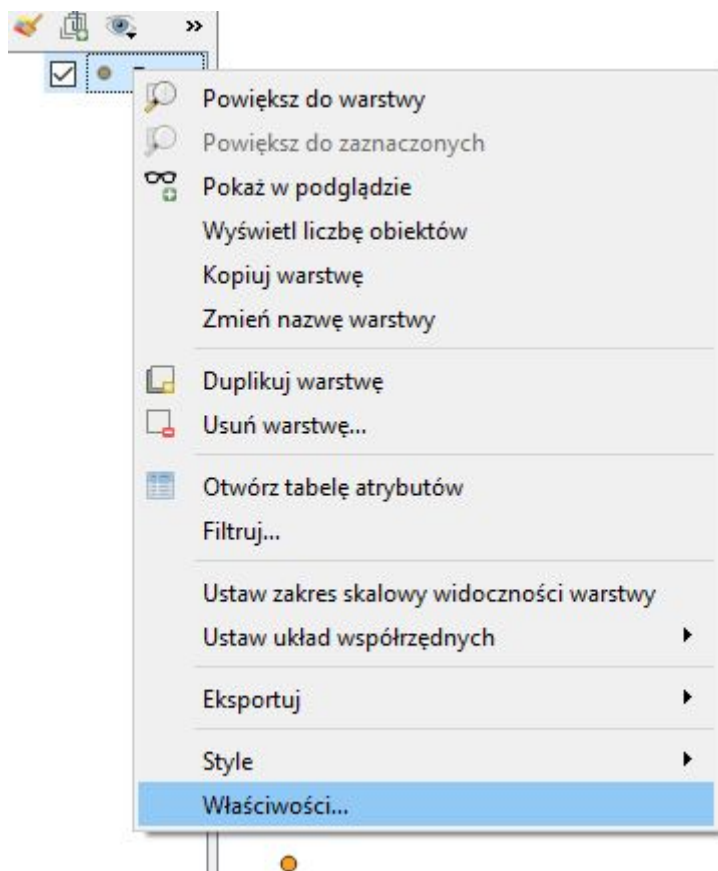


W tabeli widzimy wszystkie obiekty z przypisanymi atrybutami. Powinniśmy znać je z pliku GML.

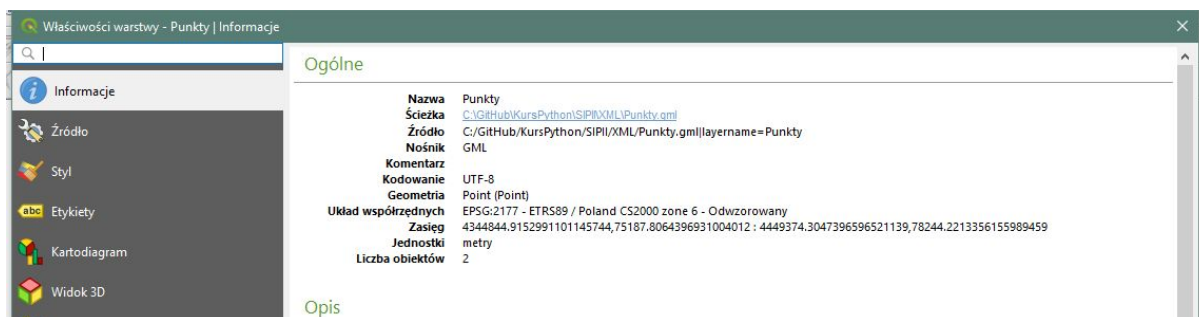
Punkty :: Liczba obiektów: 2, odfiltrowanych: 2, zaznaczonych: 0

	fid	id	Nazwa	Data_pomia	Wysokosc
1	Punkty.0	1	P1		123.320
2	Punkty.1	2	P2		123.333

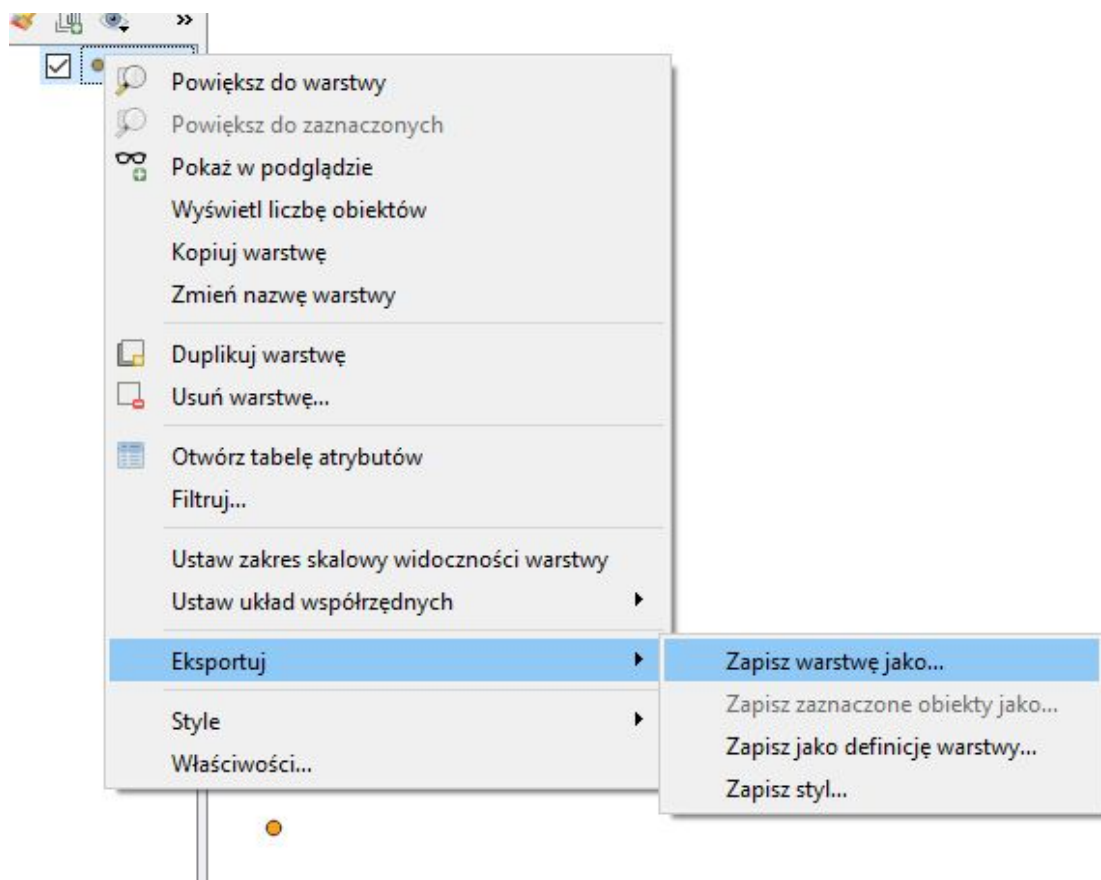
Możemy zamknąć tabelę atrybutów i przejść jeszcze do właściwości warstwy:



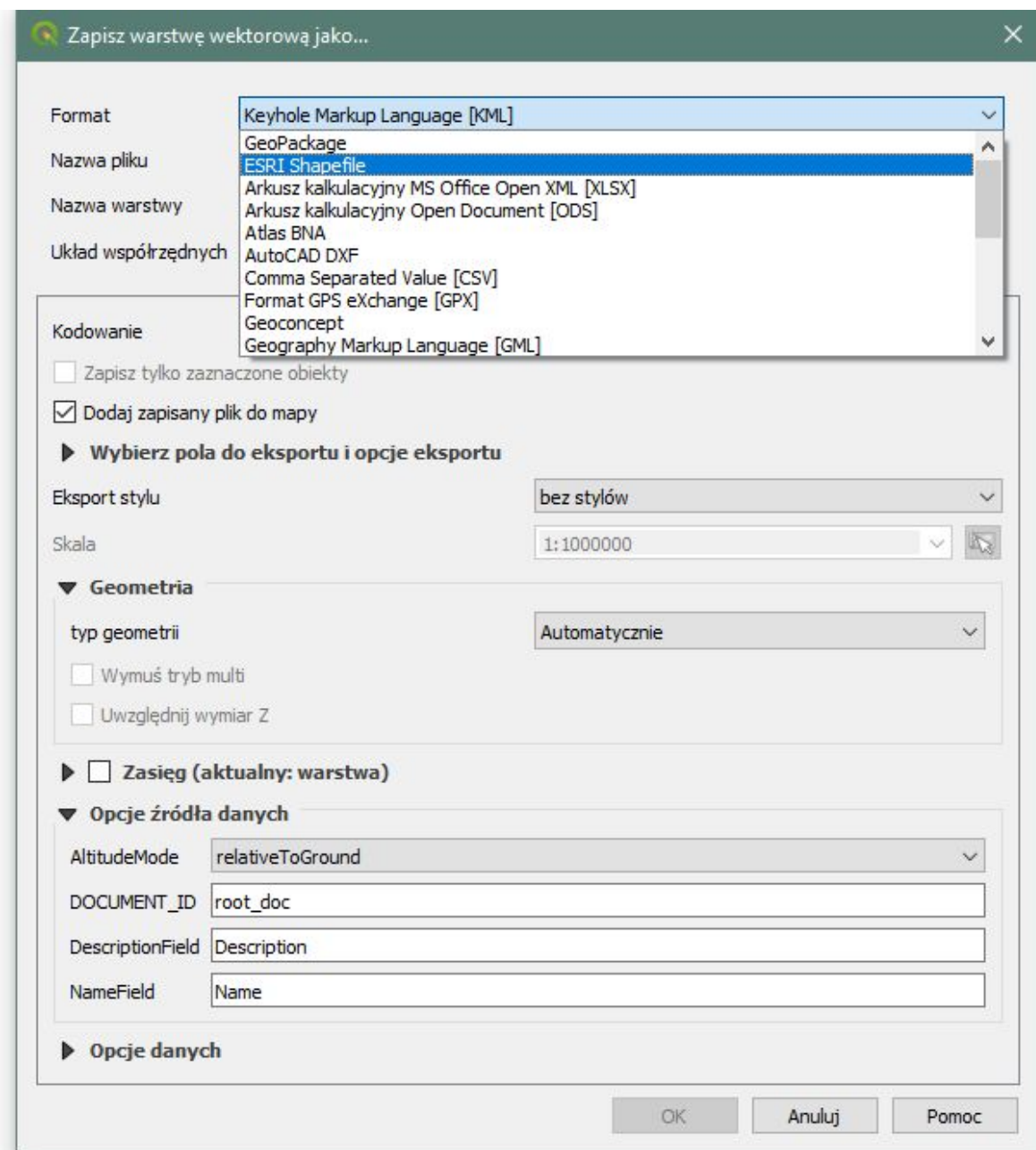
Tam w zakładce Informacje znajdziemy metadane warstwy, które również znamy z pliku GML.



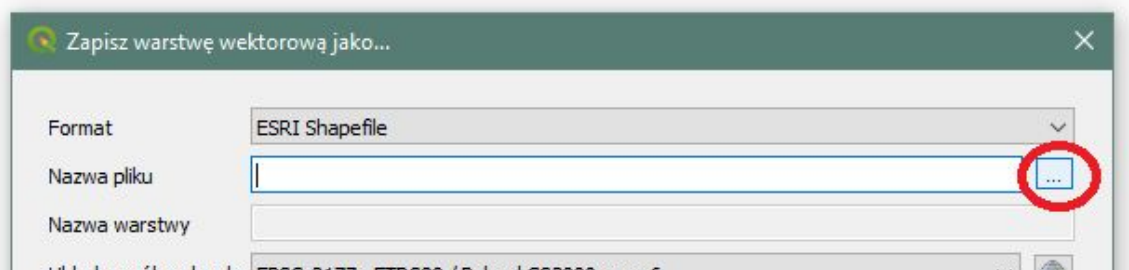
Zapiszmy jeszcze warstwę do innego formatu. W tym celu klikamy prawym na warstwę i wybieramy opcje jak poniżej:



W górnym pasku wybieramy jeden z formatów, na przykład ESRI Shapefile:



Wybieramy miejsce zapisu pliku poprzez kliknięcie w zaznaczony przycisk:



Nadajemy nazwę warstwie, wybieramy układ współrzędnych (można zostawić) i klikamy OK. Plik w innym formacie zapisze się w podanej lokalizacji.

Zadania:

1. Spróbuj określić, które informacje w pliku GML są redundantne i mogłyby zostać zapisane w inny sposób, tak by zmniejszyć jego rozmiar.
2. Na podstawie pliku "Punkty.xsd", wykorzystując swoją znajomość schematów XSD określ **wszystkie** ograniczenia, które schemat narzuca na ten plik.