

# **Dlaczego GML ?**

**Gdańsk 13.03.2017r.**

**Karol Stachura**

# Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

## **Zanim o GML najpierw o XML**

### **Dlaczego stosuje się pliki XML:**

- Tekstowe
- Samoopisujące się
- Elastyczne
- Łatwe do zmiany bez zaawansowanego oprogramowania
- Posiadające hierarchiczną strukturę
- Dostosowane do opisu złożonej rzeczywistości

Format danych powszechnie stosowany w informatyce.

# Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

```
<?xml version="1.0" encoding="utf-8"?>
<dzialka>
  <budynek id="123">
    <numer>56</numer>
    <powierzchnia>234.50</powierzchnia>
    <lokale>
      <lokal>
        <numer>1</numer>
        <powierzchnia>45.87</powierzchnia>
      </lokal>
      <lokal>
        <numer>2</numer>
        <powierzchnia>43.23</powierzchnia>
      </lokal>
    </lokale>
  </budynek>
  <drzewo>
    <gatunek>wierzba</gatunek>
    <wiek>34</wiek>
  </drzewo>
  <studnia>
    <glebokosc>15</glebokosc>
    <material>beton</material>
  </studnia>
</dzialka>
```

# Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

## **Struktura i zasady tworzenia plików XML**

- Plik xml jest plikiem tekstowym, opisującym dane, który może zawierać dowolną strukturę zagnieżdżeń znaczników pod warunkiem zachowania poniższych zasad.

Zasady tworzenia:

- musi mieć nagłówek z kodowaniem (identyczne kodowanie muszą mieć dane)
- musi posiadać jeden znacznik główny w którym zawierają się pozostałe znaczniki
- **Znaczniki muszą zaczynać się od litery lub podkreślenia**
- Znaczniki nie mogą zawierać białych znaków
- Znaczniki mogą składać się z liter, cyfr, podkreślenia kresek
- Nie powinno się stosować w znacznikach polskich znaków
- Znacznik otwierający i zamykający jest identyczny włącznie z wielkością znaków.

# Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

## **Schematy XSD**

Schematy danych XSD są plikami o rozszerzeniu xsd służącymi do sprawdzania poprawności informatycznej struktury, zgodności nazw i występowania elementów, typu danych, poprawności zagnieżdżenia plików xml.

Reguły zawarte w schematach danych mogą stanowić podstawę do tworzenia nowych schematów poprzez ich **import** do schematu który ma ich docelowo używać.

Schematy mogą być ładowane z internetu lub lokalnie. Ważne jest zachowanie prawidłowych ścieżek wewnątrz pliku xml i nie zmienianie ścieżki do pliku schematu do którego się odwołujemy.

## Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

### **Plik GML:**

Jest jednym z typów plików XML używanym do przechowywania informacji geograficznej. Schemat GML zawiera znaczniki, które mogą być zagnieżdżone w obiekcie tak aby w sposób standaryzowany dostarczył informacji o obiekcie przestrzennym, którą mogą czytać programy to obsługujące.

Pliki GML są natywnymi plikami przechowującymi informację geograficzną według normy ISO 19136 i standardów OGC.

Pliki GML o prawidłowej strukturze są czytane przez wiele programów „pudełkowych” typu GIS, ETL, bazy danych. Muszą spełniać zgodnie ze specyfikacją szereg warunków np. odwoływać się do przestrzeni nazw gml i mieć prawidłowo zagnieżdżone znaczniki i atrybuty wymagane do ich poprawnego wczytania.

# Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

## Pliki GML:

```
<?xml version="1.0" encoding="utf-8"?>
<gml:FeatureCollection xsi:schemaLocation="urn:gugik:specyfikacje:gmlas:bazaD
] <gml:featureMember> 1 wiersz w tabeli atrybutów GIS
] <gml:featureMember> 1 wiersz w tabeli atrybutów GIS
] <gml:featureMember> 1 wiersz w tabeli atrybutów GIS
] <gml:featureMember> 1 wiersz w tabeli atrybutów GIS
] <gml:featureMember> 1 wiersz w tabeli atrybutów GIS
] <gml:featureMember>
]   <ot:OT_ADJA_A gml:id="OT_ADJA_A.6">
]     <ot:idIIP>
]       <bt:BT_Identyfikator>
]         <bt:lokalnyId>FF63B765-3B00-8A27-E043-CA2BA8C0D374</bt:lokalnyId>
]         <bt:przestrzenNazw>PL.PZGiK.336.BDOT10k</bt:przestrzenNazw>
]         <bt:wersjaId>2015-11-12T00:00:00</bt:wersjaId>
]       </bt:BT_Identyfikator>
]     </ot:idIIP>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<gml:FeatureCollection xmlns:gml="http://www.opengis.net/gml/3.2" gml:id="geoinformatyka"
<gml:featureMembers>
+ <mua:AD_JednostkaAdministracyjna gml:id="ID JA 0"> 1 wiersz w tabeli atrybutów GIS
+ <mua:AD_JednostkaAdministracyjna gml:id="ID JA 32"> 1 wiersz w tabeli atrybutów GIS
+ <mua:AD_JednostkaAdministracyjna gml:id="ID JA 3215"> 1 wiersz w tabeli atrybutów GIS
+ <mua:AD_JednostkaAdministracyjna gml:id="ID JA 321503"> 1 wiersz w tabeli atrybutów GIS
</gml:featureMembers>
</gml:FeatureCollection>
```

## Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

Relacje między danymi w plikach gml można podzielić na referencje do obiektów **zewnętrznych występujących do innych rejestrów poza danymi, które wczytujemy** np. referencja do PRNG lub TERYT lub **wewnętrzne** czyli do danych w obrębie naszego zbioru danych np. referencja do słownika lub innego elementu tego samego zbioru.

Połączenia wewnętrzne wykonywane są zazwyczaj za pomocą xlink a zewnętrzne za pomocą dedykowanym im elementom gdzie klucz do połączenia jest w wartości znacznika. W różnych danych połączenia zewnętrzne są różnie zapisywane.



## Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

### **Elementy łączenia wewnętrznego – notacja xlink.**

Niestety na poziomie różnych danych, które można otrzymać z państwowego zasobu geodezyjnego i kartograficznego w Polsce xlinki w zależności od projektu wyglądają różnie:

- **emuia adresy** – wiązanie do elementu w tym samym pliku xlink:href="[gml:id]"
- **bdot, bdo** – wiązanie do elementu w tym samym pliku xlink:href="#[gml:id]" to co w emuia jest wiązaniem do elementu, w bdo jest wywołaniem całego pliku np. xlink:href="Uzytkownik12.xml"
- **PRG adresy** - wiązanie do elementu w tym samym pliku xlink:href="http://geoportal.gov.pl/PZGIK/dane/PL.ZIPIN.2837.EMUiA/30000000000031082267" - element nie istniejący w Internecie.
- **EGiB** – wiązanie do elementu w tym samym pliku xlink:href="urn:pzgik:id:[przestrzenNazw]:[gml:id]"

## Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

Walidacja danych XML/GML to automatyczne sprawdzenie czy dane są zgodne ze schematem czyli posiadają dozwolone zagnieżdżenia znaczników, wartości, atrybuty oraz czy są prawidłowe pod względem struktury wewnętrznej.

Walidacja danych ze schematem powinna być pierwszą czynnością po otrzymaniu danych od zewnętrznych dostawców. Błędy walidacji oznaczają, że dane nie są zgodne ze schematem i zawierają błędy.

Walidacja nie zawsze do końca oznacza, że zawartość pliku jest poprawna. Może wystąpić sytuacja że ciąg współrzędnych będący wartością tekstową, który powinien być zgodny ze specyfikacją GML taki nie jest, walidacja nie sprawdza też np. czy odwołania Xlink mają odpowiedniki do których się odnoszą nie sprawdzane są również wartości referencji zewnętrznej oraz np. powiązania przestrzenne między obiektami.

## Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

Pliki XSD czyli schematy danych zawierają w sobie znaczącą większość informacji niezbędnej do prawidłowego wczytania danych z pliku XML i nadania w bazie lub systemie GIS odpowiedniej struktury. Czyli każda kolumna w zależności od definicji w schemacie otrzyma odpowiedni typ np. Date, Time lub wartość liczbową czy tekstową.

Wczytywane w niektórych programach są też relacje do innych obiektów, które występują hierarchicznie. Na przykład w przypadku budynku i funkcji szczegółowej, których może być w budynku kilka - powinna być ona wczytywana jako relacja do obiektu.

Walidacja schematem XSD jest kluczowa aby móc wczytać dane do bardziej zaawansowanych programów. Brak walidacji w zależności od rodzaju błędu może powodować wyjątki. Program nie będzie wiedział jaka jest struktura danych i będzie to określał na podstawie danych zawartych w pliku XML. Może to powodować poważne problemy.

## Dlaczego GML ? - Spotkanie SGP, Gdańsk 13.03.2017r.

### **Dlaczego są takie problemy z GML ?**

- Brak ogólnie dostępnych prawidłowych schematów – na stronie GUGiK schematy XSD są w znaczącej części nie do użycia.
- Zmiany schematów przez GUGiK
- Niezgodność danych ze schematami
- Błędnie zaprojektowane struktury w schematach danych
- Błędnie wytworzone przez producentów dane GML
- Błędy techniczne przy obsłudze danych
- Wydajność - powiatowe bazy potrafią sobie nie radzić z wydaniem np. BDOT500

**GML jest formatem wymiany – jeśli dane są błędnie zagnieżdżone nie będą czytane. Nie jest to problem formatu GML, ale zdefiniowania w nim struktury danych przez projektanta schematu bez testowania wyjściowych GML.**

**Dziękuję za uwagę**