

Part 4: GML - Geography Markup Language Standards in Practice

In previous articles I described how to model a dataset in such a way that the contents are well defined (data model) and the dataset can be found and used (metadata). But what is the best way to exchange a dataset with other interested parties?

By Huibert-Jan Lekkerkerk

GML – Geography Markup Language
 What it is for: The exchange of geographic information in a structured way
 Relevant standards:
 - ISO 19136: Geographic Markup Language
 - OGC GML 3.1.1 Specification (the ISO standard is derived from this)
 Technical implementation:
 - XML scheme of the W3C
 Legal basis: The format is currently being researched as the standard exchange format within INSPIRE



Detail of the XSD scheme generated from the waterpart UML data model (see the GeoInformatics website, www.geoinformatics.com, for the entire file)

There are a number of formats available for the exchange of geographic data. Most formats have been developed with a specific software package and application in mind. Probably the best known example of such a geographic exchange format is ESRI shape files.

Open or closed?

One of the main disadvantages of a format developed by a software vendor is that it is fully compatible with the software produced by that vendor but, to a certain extent, incompatible with the software of other vendors. Further, the amount of control that users can

exert on the format is usually limited. These limitations may cause practical problems. Take, for example, a GIS package that allows multiple geometries per object. If we want to exchange this information using an exchange format that only allows one geometry per object, we either have to choose which geometry to exchange, or we have to exchange two files that the receiver then has to recombine into a single dataset. Allowing multiple geometries in an exchange format, however, is not of interest to a vendor that supports only a single geometry in its own software.

The alternative is to use open standards. With an open standard, users can define the standard. With such a standard, the need to exchange certain information usually overcomes the software limitations.

The international open standard for the exchange of geographic information is GML (Geography Markup Language). This standard was developed by the OGC (Open Geospatial Consortium) based on the information needs and wishes of a large user community.

The disadvantage of using an open standard such as GML is that, since it is developed from an information standpoint, it is usually somewhat ahead of the current software. It is therefore very important that the software vendors who have to support it in their software back such a standard. This process may, in some cases, take several years.

How does it work?

GML is a daughter language of XML (Extensible Markup Language). This language has become the language for the exchange of information over the Internet. An example of another XML-like language is HTML (Hypertext

```

<geo:Waterpart>
  <geo:identification>NL.GM.2.AGWK</geo:identification>
  <geo:name>AG Wildervankkanaal</geo:name>
  <geo:typeWater>canal</geo:typeWater>
- <geo:valueType>
- <geo:WaterValueType>
  <geo:value>107.93714361775</geo:value>
  <geo:valueType>size; length (m)</geo:valueType>
</geo:WaterValueType>
</geo:valueType>
- <geo:LineGeometry>
  - <gml:LineString srsName="urn:opengis:crs:EPSG::28992">
    <gml:coordinates>207514.725,610739.303 207522.672,610740.636
    207621.653,610753.842</gml:coordinates>
  </gml:LineString>
</geo:LineGeometry>
</geo:Waterpart>
<geo:Waterpart>
  <geo:identification>NL.GM.2.RFC DOMAIN 20131</geo:identification>
  <geo:typeWater>other: ecological-watercourse</geo:typeWater>
- <geo:valueType>
- <geo:WaterValueType>
  <geo:value>268.031685166776</geo:value>
  <geo:valueType>size; length (m)</geo:valueType>
</geo:WaterValueType>
</geo:valueType>
- <geo:LineGeometry>
  - <gml:LineString srsName="urn:opengis:crs:EPSG::28992">
    <gml:coordinates>205637.201,610450.66 205643.894,610396.188
    205668.946,610184.516</gml:coordinates>
  </gml:LineString>
</geo:LineGeometry>
</geo:Waterpart>

```

Markup Language) which is used to code web pages. In practice most users will not be confronted with either XML or HTML: our web browser translates the underlying fields into comprehensible information such as a webpage or table.

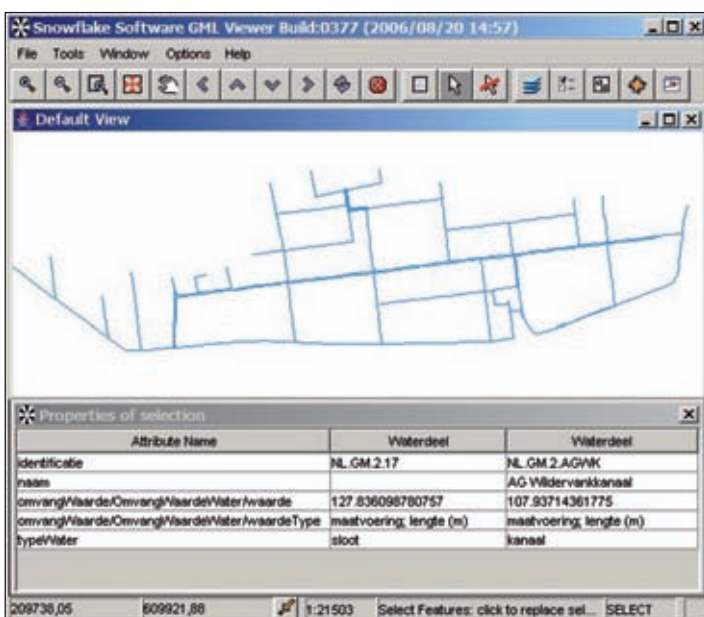
GML file

A GML file is easy for the user to read, that is, if one knows what there is to be read. The file itself is decoded using readable text (ASCII) and is not, as is the case with most other exchange formats, based on binary characters. Therefore anyone with text-editing software can change a GML file.

The structure of a GML file takes some getting used to. As with a number of formats,

the file starts with a header that states the information the computer needs in order to read the remainder of the file. The information after the header has a strong structure and looks somewhat similar to a database. The data itself is written in so-called fields. In the example these fields can be seen as text between tags (recognizable from the <>). GML allows the addition of multiple (sub) fields to a single main entry or object. This way all the information related to an object can be stored with that object, creating a clear information structure. Relations between objects can also be defined using pointers or nested structures.

GML scheme



Sample GML file as a geographic image in the Snowflake software

A GML file is easy for the user to read, that is, if one knows what there is to be read.

Example of the coding of two waterparts into a GML file (see the GeoInformatics website, www.geoinformatics.com, for the entire file)

type of data that is allowed in a particular field including the length of the field. The 'pull-down lists', the enumerations and codelists, are defined within the scheme as well. This way a self-describing format is created. Essentially all information needed to read the GML file is defined in the associated scheme.

Validation

The advantage of using the combination of a scheme and data file is that the latter can be validated for both the structure and the type of data in the fields. All mandatory fields from the scheme (and therefore the data model) have to be filled in the GML file. If they are not present the software usually gives an error message. An error message is also given if a non-existing value is selected for an enumeration field.

The software will not give an error or warning when technically correct which means impossible data is entered into a field. If, for example, the length of a watercourse is defined and someone types in 0.01 m, the file will pass the validation. A check on this type of error is not built into the current XML format and has to be checked by the software itself.

Geometry

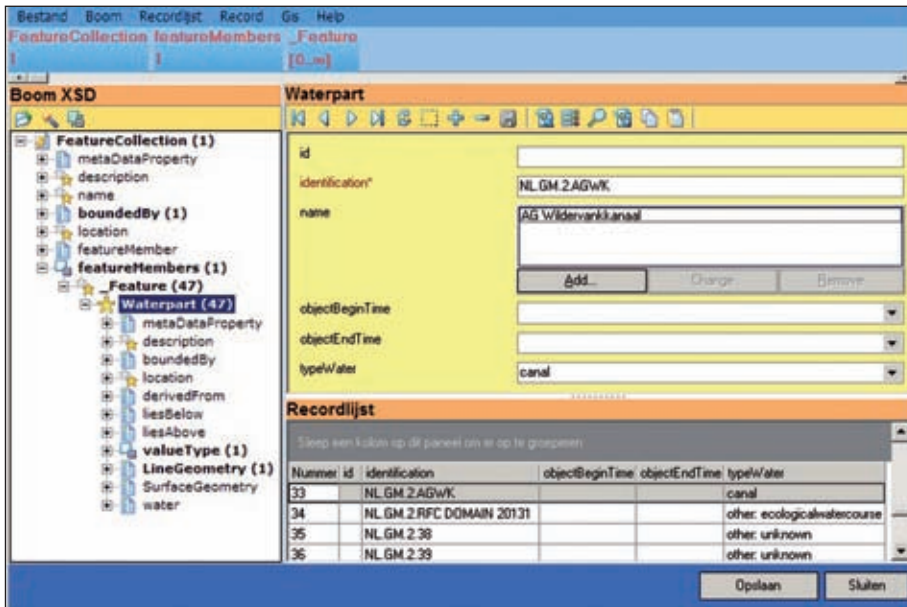
The fact that all object information is exchanged in a structured way is great, but with geographic information what really counts is the relation to the geometry of the object.

Within GML this has been solved in such a way that quite a number of geometry and topology types are available for the geographic component of the data. Examples of geometry types are, of course, point, line and surface geometries. But more complicated types such as multipoint, 3D solids and multi-surfaces are allowed. Even mixing basic geometry types into a single geometry is possible.

Profiles

GML supports more geometry types than most current software does, including the 3D geometry types that as yet are available in just a few GISs.

This is also true for the way the object information (attributes) is structured; for example,



Sample GML file as administrative data in the Genetics XML viewer

Legalization

Just as with data models, there is no legal requirement for the use of GML as an exchange format. More and more governments are, however, switching to the use of open standards for government data. At the moment the only worldwide open standard for geographic information exchange is GML. The European INSPIRE project is currently investigating the use of GML as a basis for the exchange of information, but is not yet at the point of defining this in the implementing rules. Finally, when using WFS (Web Feature Service), the use of GML is prescribed in this standard. More on WFS (and WMS) in the next article.

Huibert-Jan Lekkerkerk

(hlekkerkerk@geoinformatics.com /

h.lekkerkerk@idsw.nl) is project manager standardization at IDsw and editor-in-chief of *GeoInformatics*.

For more information: OGC: www.opengeospatial.org;

INSPIRE: www.ec-gis.org; Snowflake:

www.snowflake.co.uk; Genetics: www.genetics.nl.

GML, similar to UML, allows an attribute to be used multiple times with a single object. This is, however, not allowed in all software. To prevent this type of problem, GML has defined a number of standard profiles. These profiles limit the GML possibilities in a controlled way. The standard profiles are called

SF (simple features) and are defined on three different levels: 0, 1 and 2. SF level 0 has roughly the same possibilities as the well-known shape format. The higher the level, the more GML options are allowed for in the exchange file, and the more options the software has to support.