Jarosław Bydłosz*, Piotr Cichociński*, Izabela Basista*

The Possibilities of Geoinformation Resources Recorded in GML Accessing with Chosen GIS Software**

1. Introduction

Development work – named INSPIRE – related to implementation of a directive aimed to set the infrastructure of spatial information within European Union is now under progress. According to this directive spatial information infrastructures in EU member states should be designed to ensure linking of spatial data originated from various sources in uniform manner, thus ensuring possibility of common use by many end users and applications.

At the same time all institutions grouped under Open Geospatial Consortium (OGC) have been developing a standard Geography Markup Language (GML) for a few past years. It is designed to serve as a geographic system modeling language, also as the registry of geographic information and standard for geographic data exchange through the Internet. The objective of this paper is to demonstrate functions that the GML language may fulfill for accessibility of data included in the Polish geoinformation resources while using presently accessible GIS software.

2. INSPIRE Directive

Directive of the European Parliament and the Council [4] establishing an Infrastructure for Spatial Information within European Community (INSPIRE) was issued on March 14, 2007. This directive sets legal provisions for establishment and activities pertaining to the Infrastructure for Spatial Information within

* Faculty of Mining Surveying and Environmental Engineering, AGH University of Science and Technology, Krakow

** This work work is connected with scientific works conducted in the Department of Geomatics, AGH University of Science and Technology

the European Union. Union policy aims at the highest levels of environmental protection while considering a diversity of situations existing within various regions of the Community. In order to fulfill these requirements, it is necessary to provide appropriate coordination between the subjects that supply information and the end users, as well as facilitate possibility to combine information and knowledge originated in various sectors. First of all, the INSPIRE directive should be applied for spatial data stored by the public authorities or on their behalf.

According to this directive, the infrastructures of spatial information established in member states should be designed to provide the following:

- ensure storage, availability and maintenance of data;
- make it possible to link spatial data originated in various Community sources as well as their mutual use by numerous users and applications;
- facilitate common use of spatial data stored at one level of public institutions by other authorities;
- make the spatial data available on conditions that do not limit their wide use without a serious reason;
- easy search of accessible spatial data and easy access for a given purpose as well as simple conditions of its application.

According to INSPIRE directive, the member states should create a network encompassing a number of services, such as: searching, browsing, downloading, processing as well as other services pertaining to initiation of spatial data. The member states should submit their first report to the hands of European Commission before May 15, 2010.

Such report is to be submitted every three years and should contain the following information:

- methods for coordination of public sector suppliers, data set and spatial service users as well as intermediate authorities;
- contribution of public sector suppliers or third parties to functioning and coordination of spatial information infrastructure;
- information pertaining to use of spatial information infrastructure;
- agreements contracted between public institutions or subjects pertaining to mutual use of the data;
- costs and benefits resulting from implementation of this directive.

3. GML

The OGC consortium (Open Geospatial Consortium – the organization grouping together the leading producers of GIS software among the other institutions) established in 2001 that the GML language (Geography Markup Language) will be

used as a standard for coding, distribution and acquisition of geographical information data. Its specifications were approved by the International Standard Association (ISO) in 2007 in form of international standard [7], and then approved by the Polish Standardization Committee (PKN) to be added to a set of the Polish standards. GML is the implementation of the XML language (eXtensible Markup Language) [1].

XML is simply a formal markup language that can be used to exchange information between computer systems. Its main application is to register structured documentation. The XML language may be used in relatively easy way to create and supervise hierarchical organizational data structures. Those structures are met in geographical applications quite often. XML does not determine semantics or sets of markers.

Truly speaking, XML is a language describing the data, or more clearly, a metalanguage used to express data description languages (markup languages). Since it is also extendable, then practically speaking, any user who wants to write the data may do so according to his needs. However, it is necessary to add schema files to the file containing data, thus allowing to correctly read (decode) that data. Schema is basically a model describing the structure of information. It is an expression originating from the world of databases, where the data structures are described by relational tables. Therefore, schema describes a model for a whole class of documents within XML context. This model describes possible setting of markers and text in a structurally correct document. One can also look at schema in terms of an agreement of common vocabulary for a particular purpose which requires exchange of documents.

Models pertaining to schemas are often described by means of design principles. The principle defines what can appear within a given context. Essentially, there are two types of principles: principles of content model describe a sequence of appearing elements, while data type principles define correct data units. Authors of documents may define the structure and permissible types of data by using schemas. Processing software programs may then check consistency of documents which are expected to be instances of a given schema. In conclusion, GML is just such a schema designated to collect and transfer geographic information. The term geographic information is considered to be a combination of both geometry as well as attributes (properties) of geographic objects. GML specification defines the mechanisms and a syntax used by GML to store the geographic information in XML format. Primary idea to create GML language was to apply XML language for storage of geographic information in order to exchange it without a loss between various systems, regardless of hardware and software system platforms as well as regardless of the type and technology of geoinformation system.

4. GML vs. Transposition and Implementation of INSPIRE Directive in Poland

During the ongoing second phase of the INSPIRE program development, the basic task for the member states, including Poland, is to perform a transposition of the INSPIRE directive, that is to introduce its regulations to the country's legal systems [3]. The project of an act pertaining to spatial information infrastructure [10] was created in consistence with an accepted transposition method [2]. Among the other things it will also include the appointment of a leading institution for each of 34 spatial data topics. The mentioned leading institution could be a minister or the central government administration office. The country's Surveyor General is a leading institution within 15 topics pertaining to the data which is maintained fully, or in part, by the surveying and cartographic services.

During the first stage of work pertaining to the development of implementing provisions with regard to interoperability, one of the first four documents created was the *Guidelines for the Encoding of Spatial Data* [3]. It contained a description of geoinformation method which would enable transfer between systems used by the Member States. GML was approved as a preferred (default) language.

The following solutions will be expected in the content of legal regulations and guidelines being designed and connected to transposition of the INSPIRE directive to Polish legal system within the scope pertaining to GML [9]:

- in connection to current amendments to the act pertaining to Surveying and Cartographic Law [11], it is planned to announce new implementing provisions, assuming use of ISO standards, modeling and exchange standards based on XML and GML;
- in case of topographic objects showing details similar to the basic map – it is expected to simplify the structures and provide access to databases in GML format (or other formats, depending on needs of the data end users);
- in reference to basic maps as well as surveying maps of area utilities – it is expected to simplify the GESUT (Infrastructure Cadastre) structure to enable its access in GML format (it will also apply to other structures – i.e. shapefile).

5. Internet Map Servers

One of the issues touched upon in the above mentioned INSPIRE directive is the ease of access to dispersed national or European spatial data resources. The directive's assumptions are obligating us to establish a National Infrastructure of Spatial Data using the Internet, and more specifically, through special services im-

plemented on servers. According to the directive, those services should make it accessible for the user to browse and download their contents as well as spatial data sets. Moreover, they should also take into consideration all relevant user requirements, be easy to work with and accessible for the public use [4].

These efforts should be combined with a takeover of the European Union recommended worldwide technical standards pertaining to spatial information infrastructures developed by the ISO/TC211 Technical Committee and OGC. Use of these standards will allow achieving interoperability between components of the established infrastructure. In accordance with those standards, browsing through spatial data should be done by means of WMS (Web Map Service), while data acquisitions are enabled through WFS (Web Feature Service) specification.

WMS standard provides interface for transmission of raster image data converted to legible form by internet browsers upon downloading from database. It is already used within the GEOPORTAL.GOV.PL project. Geoportal is an infrastructure of nodes contained within the National Infrastructure of Spatial Information (KIIP), interacting with each other and rendering services of searching, making available and analyzing the spatial data in consistence with the INSPIRE directive. Developments and spatial data featuring cadasters, orthophoto air and satellite maps, sozologic (environmental) maps or data pertaining to administrative borders of national area divisions are made available already, among the other items, on Geoportal websites. That information can be accessed in a form of forwarding or guiding to internal data (pertaining to any services with spatial data stored in the system). However, one should realize that the WMS standard does not provide possibility for geospatial data transfer, but only for a graphical representation of those data.

WFS servers will be used to provide the data in vector formats. This standard is based on GML language. WFS service utilizes six operators. These are:

- *GetCapabilities* – allowing for the map server to describe its possibilities and stored data;
- *GetFeature* – allowing to select relevant objects;
- *DescribeFeatureType* – delivering a description of XML schema to one or more object features; therefore it is possible to find information about data attributes and types;
- *Transaction* (optional) – making it possible to perform all operations which allow to identify the data;
- *LockFeature* (optional) – enabling to create data locking in order to provide possibility of multiple access to various objects;
- *GetFeatureWithLock* (optional) – is a combination of *GetFeature* and *LockFeature* [8].

It is easy to notice that the use of GML to make spatial data available through WFS service will allow formulating a number of questions containing text and spatial criteria, thus providing an easy access to attributes as well as the spatial data itself. In response to the inquiry submitted to server, the user receives a file with.xml extension, containing spatial data selected in accordance with the question. A sample question and WFS server response is presented below.

Inquiry:

```
http://www.bsc-eoc.org/cgi-bin/bsc_ows.asp?version=1.0.0&
service=WFS&request=GetFeature&TypeName=OBBA_REGION
```

Listing 1. Request to the server concerning administrative unit borders from the Bird Studies Canada OBBA (*Ontario Breeding Bird Atlas*).

Server response:

```
<?xml version='1.0' encoding="ISO-8859-1" ?>
<wfs:FeatureCollection
    xmlns:bsc="http://www.bsc-eoc.org/bsc"
    xmlns:wfs="http://www.opengis.net/wfs"
    xmlns:gml="http://www.opengis.net/gml"
    xmlns:ogc="http://www.opengis.net/ogc"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.opengis.net/wfs
    http://schemas.opengeospatial.net//wfs/1.0.0/WFS-basic.xsd
    http://www.bsc-eoc.org/bsc http://www.bsc-eoc.org/cgi-
    bin/bsc_ows.asp?SERVICE=WFS&amp;VERSION=1.0.0&amp;
    REQUEST=DescribeFeatureType&amp;TYPENAME=OBBA_REGION&amp;
    OUTPUTFORMAT=XMLSCHEMA">
        <gml:boundedBy>
            <gml:Box srsName="EPSG:4326">
                 <gml:coordinates>-95.292625 ...</gml:coordinates>
            </gml:Box>
        </gml:boundedBy>
     <gml:featureMember>
        <bsc:OBBA_REGION>
           <gml:boundedBy>
              <gml:Box srsName="EPSG:4326">
                   <gml:coordinates>-93.153435 ...</gml:coordinates>
              </gml:Box>
           </gml:boundedBy>
           <bsc:msGeometry>
           <gml:Polygon srsName="EPSG:4326">
              <gml:outerBoundaryIs>
                 <gml:LinearRing>
```

```
          <gml:coordinates>-93.153099,54.149952 -93.153435,54.23982 ...
                </gml:LinearRing>
             </gml:outerBoundaryIs>
           </gml:Polygon>
           </bsc:msGeometry>
           <bsc:REGION>43</bsc:REGION>
          <bsc:LINK>http://www.birdsontario.org/atlas/datasummaries.jsp?
          extent=Rg&amp;summtype=SpList&amp;year=allyrs&amp;region1=43
          </bsc:LINK>
        </bsc:OBBA_REGION>
      </gml:featureMember>
...
```

Listing 2. The server answer – the XML file with with OBBA_REGION objects containing geometry and descriptive attributes.

This inquiry pertains to a selection of data concerning OBBA (*Ontario Breeding Bird Atlas*) administrative unit borders from the Bird Studies Canada server (BSC, http://www.bsc-eoc.org/cgi-bin/bsc_ows.asp). As a response, we are getting XML file with OBBA_REGION objects containing geometry (<bsc:msGeometry>) as well as descriptive attributes (<bsc:REGION>,<bsc:LINK>).

6. GML File Import

In reference to spatial data pertaining to the Polish territory, the authors were unable to find a WFS server making such data available. But there are foreign WFS servers that enable to download vector data without any problems. In case of Geoportal, unavailability of such data is caused, among the other reasons, by a lengthy WFS format, resulting from geometry notation in a text file (GML) and leading to a quick increase of the file volume as the presentation of larger area progresses. Besides, the data transmitted by WFS must be processed by a client or WWW server which presents that data and that, in turn, is combined with time consuming operation of converting vectors to rasters. Therefore, the WFS format is only used internally by the Geoportal server. Its main task is to perform those services which are not accessible for WMS that is searching and marking of addresses. WFS format will also be utilized by Geoportal to sales of numerical data in vector formats [12].

Due to inaccessibility of spatial data originating from WFS servers, the authors decided to use data from the Topographic Database (TBD) which will be accessible through the internet in the future. That data was written in GML language in accordance with a specification determined in technical guidelines of the Topographic Database (TBD) [6]. The set of data for one sheet consists of several files,

whereas is file contains information originated from a single subject layer (buildings, roads, groundwater flows, etc.). Each GML file should be accompanied by a schema file, since only in that case it is possible to read correctly its structure and content. Therefore, in case of TDB, this will be a set of schema files. The schema files describing both topographic as well as cartographic parts of TBD are available on the website under http://www.gugik.gov.pl.

This data was used to check how well could the software available on the market cope with reading the files written in this standard. Commercial ArcGIS software as well as free Quantum GIS and OpenJUMP software were thoroughly tested. In accordance with their experience, the authors show that breaking the data into two components (essential data and schema) appears to be the main issue with GIS files written in GML format.

ArcGIS – this program has an extension – *Data Interoperability*, which provides a conversion between multiple file formats as well as allows defining its own input and output formats. It was due to this feature that a selection of relevant files containing schemata was possible. A very useful property of this tool was also a list of ongoing operations performed during the process of data reading. It allowed catching the errors, in particular those that were caused by wrong placements of GML files containing the spatial data, in relation to locations of the file sets containing schemas. The main schema file, associated with a particular object, has a number of additional supplementing schema files; therefore the program had to identify the whole set of schemata in order to read the spatial data correctly. In relation to the data file, each of those schema files has a notation of the access path to the associated schema file. As a result, when reading the main schema the program can localize next necessary schemata.

The program coped very well, practically without any problem, with reading of GML file containing one spatial object. In case of a set of objects (several files with data), it was necessary to indicate the whole set of objects to be read, together with one paramount schema file containing access paths to the catalogue with main object schemas.

OpenJUMP – one of more interesting free programs dealing with GML format data import. The program allows to read GML data stored according to its internal schema JUMP GML (.jml) as well as files FME GML (.gml, .xml, .fme). Moreover, there is a possibility to point out to schema file, or rather a template, since its structure is totally different than that of a standard schema, but consistent with one available in the OpenJUMP program documentation.

For one of the TBD files with a record of ADGM_A object:

```
<Dane>
    <ADGM_A>
        <X_KOD_TBD>ADPA01</X_KOD_TBD>
```

```
<X_AKTUALNOSC_G>2006-02-28</X_AKTUALNOSC_G>
<X_AKTUALNOSC_A>2006-02-28</X_AKTUALNOSC_A>
<X_KAT_DOKL_GEOM>2</X_KAT_DOKL_GEOM>
<X_ZRODLO_DANYCH_G>PRG</X_ZRODLO_DANYCH_G>
<X_ZRODLO_DANYCH_A>PRG</X_ZRODLO_DANYCH_A>
<X_KAT_ISTNIENIA>1</X_KAT_ISTNIENIA>
<X_RODZAJ_REPR_GEOM>ZU</X_RODZAJ_REPR_GEOM>
<ID>1</ID>
<ID_TERYT>3064011</ID_TERYT>
<NAZWA>M. Poznan</NAZWA>
<ID_POWIATU>3064</ID_POWIATU>
<NAZWA_POWIATU>Poznan</NAZWA_POWIATU>
<OBSZAR>
     <gml:Polygon>
          <gml:outerBoundaryIs>
               <gml:LinearRing>
                    <gml:coord>
                         <gml:X>364025.18</gml:X>
                         <gml:Y>507520.82</gml:Y>
                    </gml:coord>
                    ...
```

Listing 3. TBD file with a record of ADGM_A object.

The file template defining various ADGM_A object attributes will look as shown below.

```
<?xml version='1.0' encoding='UTF-8'?>
     <JCSGMLInputTemplate>
         <CollectionElement>Dane</CollectionElement>
         <FeatureElement>ADGM_A</FeatureElement>
         <GeometryElement>OBSZAR</GeometryElement>
         <ColumnDefinitions>
             <column>
                 <name>X_KOD_TBD</name>
                 <type>STRING</type>
                 <valueElement elementName="X_KOD_TBD"/>
                 <valueLocation position="body"/>
             </column>
             ...
         </ColumnDefinitions>
     </JCSGMLInputTemplate>
```

Listing 4. The file template defining various ADGM_A object attributes.

Unfortunately, a creation of such schema file is not easy; since it requires a great deal of knowledge about the data contained in an imported GML file (in this case it was necessary to get acquainted with TBD Technical Guidelines). Moreover, despite pointing out to a relevant template, the program does not read

GML files without the geometry content (or tables containing only descriptive information about the object).

Quantum GIS – this program uses imported GML files which belong to free software group, written in C++ language of the OGR *Simple Features Library* [5], which provides possibility of readout (and in some cases also writing) of vector data sets recorded in various formats, such as: ESRI Shapefile, S-57, SDTS, PostGIS, Oracle Spatial as well as Mapinfo mid/mif and TAB.

As opposed to majority of other converters, this tool does not use the object class definition contained within the schema file. Instead, it attempts to detect it automatically, by reviewing GML file contents and searching the typical objects from GML name area in order to determine its structure. Unfortunately, the data from TBD could not be read with this program. The program displayed a message saying that the file to be opened „is not a correct and recognizable data set".

Varying results with TBD file imports motivated the authors to attempt entering other files. They used samples of GML files accessible in the internet and originating from the following topographic databases: Ordnance Survey MasterMap (Great Britain), AAA-NAS (Germany) and TOP10NL (The Netherlands).

ArcGIS coped quite well with entering those data sets, without starting the extension *Data Interoperability*, however, one could suspect that it used the profiles of two entered data sources (OS Master Map and TOP10NL), written into this extension.. There is no assurance, how AAA-NAS data was entered, however, one could suspect that a support for this format is (as is also for the remaining two formats) embedded in the extensions by the manufacturer.

However, attempts to import it with Quantum GIS rendered mixed results. The least problems appeared to be with OS Master Map. All the objects, regardless of the type of geometry and set of associated attributes, presented in one layer, were entered and presented during the first test iteration, but this resulted in ignoring of those attributes. Therefore, a specific OGR library feature was used. When the GML file is opened by this library for the first time, it is fully scanned in order to determine the types of entered objects and their attributes. Additionally, the contents of various fields with data are analyzed for possibility of determination of types. The information about GML file structure received this way is then written into a file with.gfs extension. Subsequent attempts to open the same GML file are then based on the same.gfs file, and that speeds up an access to the data.

The edition of a relevant.gfs file was performed, leaving a description of only one class of objects:

```
<GMLFeatureClassList>
  <GMLFeatureClass>
    <Name>AX_Fahrwegachse</Name>
    <ElementPath>AX_Fahrwegachse</ElementPath>
```

```
        <DatasetSpecificInfo>
           <FeatureCount>1350</FeatureCount>
        </DatasetSpecificInfo>
        <PropertyDefn>
           <Name>breiteDesVerkehrsweges</Name>
           <ElementPath>breiteDesVerkehrsweges</ElementPath>
           <Type>Integer</Type>
        </PropertyDefn>
        <PropertyDefn>
           <Name>funktion</Name>
           <ElementPath>funktion</ElementPath>
           <Type>Integer</Type>
        </PropertyDefn>
        <PropertyDefn>
           <Name>hatDirektUnten</Name>
           <ElementPath>hatDirektUnten</ElementPath>
           <Type>Untyped</Type>
        </PropertyDefn>
     </GMLFeatureClass>
  </GMLFeatureClassList>
```

Listing 5. The part of.gfs file containing information about the structure of GML file.

This allowed entering the selected class including a full list of attributes. Similar action was done with AAA-NAS set; however, in case of the whole set as well as its part, no objects were entered, although its structure was recognized correctly. On the other hand, an attempt to open TOP10NL set was terminated with a display of the message saying that the file to be opened "is not a correct and recognizable data set".

7. Conclusions

GML language has formally been in use for a few years, but did not win an adequate place among the users of Geographic Information Systems. It is not surprising. It appears to be self explanatory based on the above mentioned results of the GIS program analysis pertaining to possibilities of entering the spatial data in GML format. Although more and more programs are equipped with tools to enter such files, it is not that easy in practical use. In order to correctly enter the data, users must know at least the bases of the GML language. However, it should be admitted that in many cases it is possible to enter the data written in this particular way. In order to facilitate this process it would be necessary to create the extended application to enable correct entering the GML files together with

schemata. Truly speaking, a possibility of entering the spatial data depending on specific data and user needs is a great advantage of the GML language format. The authors hope that the development of GIS programs will also enable better migration of data written in GML standard. So far it appears that we are not fully prepared for that not only within the field of legislation but also from the point of view of technical implementation of INSPIRE directive guidelines.

References

[1] Cichociński P.: *Język XML i jego implementacje dla danych przestrzennych*. XI Konferencja Naukowo-Techniczna "Systemy Informacji Przestrzennej", Warszawa 2001.

[2] Gaździcki J.: *Aktualne problemy polityki geoinformacyjnej państwa w kontekście INSPIRE*. Roczniki Geomatyki 2007, t. V, z. 6.

[3] Gaździcki J.: *Druga faza programu prac INSPIRE: stan w Europie i wyzwania dla Polski*. V Sympozjum "Krakowskie spotkania z INSPIRE", Kraków 2009.

[4] *Directive 2007/2/Ec of The European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. European Parliament and Council, 2007.

[5] *OGR Simple Feature Library*. GDAL, 2009 [on-line]. Available at: http://www.gdal.org/ogr/index.html.

[6] Wytyczne techniczne: *Baza Danych Topograficznych*. Główny Urząd Geodezji i Kartografii, Warszawa 2003.

[7] ISO 19136: *Geographic information – Geography Markup Language (GML)*. 2007.

[8] OGC *Web Feature Service Implementation Specification*. 2005.

[9] Orlińska J.: *Implementacja INSPIRE w zakresie geodezji i kartografii*. V Sympozjum "Krakowskie spotkania z INSPIRE", Kraków 2009.

[10] *Ustawa z dnia... o infrastrukturze informacji przestrzennej. Projekt z dnia 03.11.2009 r.* [on-line]. Available at: www.gugik.gov.pl/gugik/dw_files/1378_projekt_ustawy_inspire_30_04_2009.pdf.

[11] *Ustawa z dnia 17 maja 1989 r. Prawo geodezyjne i kartograficzne*. Dz. U. z 1989 r. Nr 30, poz. 163 z późn. zm.

[12] *Dystrybucja geometrycznych danych wielkoskalowych oparta o serwery WMS i WFS. Część pierwsza – standard WMS.* Zespół ds. Krajowej Infrastruktury Danych Przestrzennych (KIDP), 2007 [on-line]. Available at: http://www.gugik.gov.pl/gugik/dw_files/758_specyfikacja-wmswfs.pdf.