Chapter 4

Agnieszka Chojka

GIS Data Standard Needed for the Environment Protection Services According to the ISO 19100 Series of International Standards

1. Introduction

In the environment protection domain a GIS data standard was developed (Łochyński, Guzik 2009). It is called, by its authors, a common language for the environment protection services for their needs of collecting and exchanging geographic information. In 2009, in the cooperation with the Conservation of Nature Institute of Polish Academy of Science in Cracow, for commission of the Ministry of the Environment, the 3rd version of this GIS data standard was developed and submitted for public consultations that were carried out by the General Directorate for Environmental Protection.

The aim of proposed GIS data standard is among other things the realization of the INSPIRE directive regulations (Directive 2007) and the ability to collect spatial data in the units of national parks, landscape parks and their administration, the general directorate for environmental protection and the regional directorates, which are derived from stock-taking of nature protection design, as well as the ability to make the spatial central register of nature forms conservation and its regional versions.

The current documentation of this standard includes as follows:
- description of data structures for geodatabase in ArcGIS to collect data concerning nature and its conservation forms,
- dictionaries with coded attribute values,
- description of graphical symbols structures to present data via Internet.

Although the authors of above mentioned standard have done a tremendous work, their elaboration has some disadvantages. The current and full version of this standard was received from its authors (GIS Standard 2010).

The aim of this paper is not to make an analysis of the GIS data standard content, because its domain – the environment protection – is not a filed of author's research. The essence of this paper is to make an analysis of this standard from geoinformation point of view and to present its conception with the use of UML notation – the formal mean of geographic (spatial) information modelling recommended by the ISO 19100 series of International Standards in the geographic information (geoinformation/geomatics) domain, as well used in the INSPIRE data specifications.

The authors of discussed GIS data standard needed for the environment protection services state that their elaboration takes into consideration the spatial data themes listed in annexes in the INSPIRE directive (Directive 2007), but they didn't take into account requirements imposed on the data specifications for protected sites (Protected Sites 2010). The reason for this was "the specific character of INSPIRE data and the purpose of this GIS data standard preparation". The aim was not to translate the INSPIRE data specifications into Polish, but to develop guidelines that could give opportunities to create "homogeneous spatial data for problems connected with natural environment and environment protection in Poland". The authors of discussed standard propose to accommodate their standard to requirements of data specifications at the central level as a result of data transformations (that is analysis, interpretation and generalization) (GIS Standard 2010). And this offer is undoubtedly worth considering.

2. INSPIRE and UML

The INSPIRE directive leads to build the ESDI (European Spatial Data Infrastructure). It is possible to consider two views on spatial data infrastructure: data-centric view and service-centric view. The data-centric view primarily addresses concerns related to the data, which includes application schemas and metadata. The service-centric view includes the taxonomy of services, concepts of interoperability, service architecture, service catalogue and the underlying IT standards.

The data-centric view is connected with the model-driven approach, where the information is described by a formal, implementation-independent schema. Implementations for various techniques (e.g. XML file transfer, web services, relational database) and implementation environments (e.g. J2EE, .Net) can be derived from the schema in a more or less automatic way. Changes in information requirements are applied to the schema, never directly to the implementation (Fig. 1).
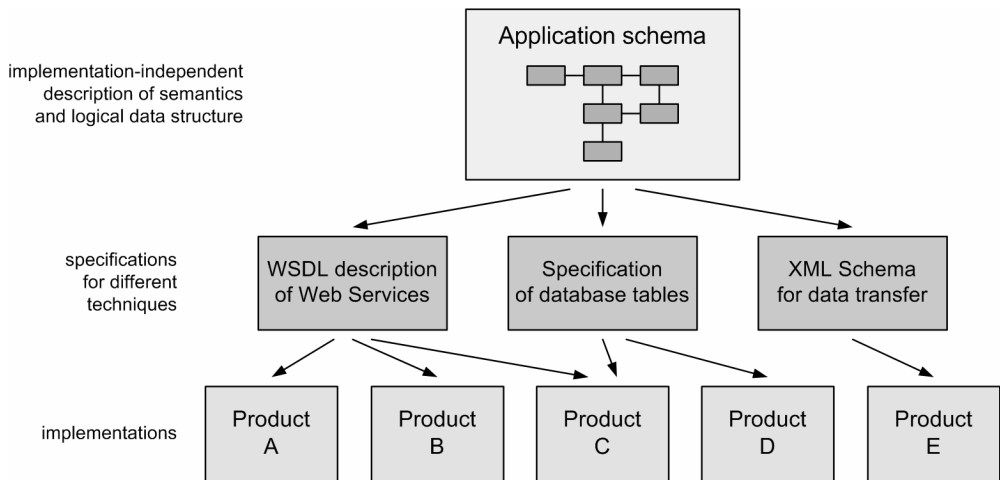
Fig. 1. The model-driven approach; source: prCEN/TR 15449 2006.

The model-driven approach follows the concepts developed in the model-driven architecture (MDA) defined by OMG (Object Management Group), which aim is to resolve problems concerning integration of information systems derived from different software developers and working on different platforms (using various technologies, e.g. various operating systems, network standards, programming languages).

According to the MDA idea, four more and more detailed models of system are distinguished:

– CIM (Computation Independent Model) – presents the requirements for the system,
– PIM (Platform Independent Model) – conceptual model – describes the system, but does not show details of its use of its platform (defines main concepts from the specific application domain),
– PSM (Platform Specific Model) – logical model – model of the same system specified by the PIM, it also specifies how that system makes use of the chosen platform,
– Implementation Model – physical model – e.g. physically working application, empty database structure.

The key role in the MDA technology plays modelling the system in UML (Unified Modeling Language), which is also a conceptual schema language recommended by the ISO 19100 series of International Standards.

3. The ISO 19100 series of International Standards

The purpose of the ISO 19100 series of International Standards in the geographic information domain is to enable interoperability between heterogeneous geographic information systems (GIS). The interoperability stands for an ability to

co-operate, therefore it is an ability to communicate, what follows achieving the data transfer between systems.

To achieve interoperability between heterogeneous systems two fundamental issues need to be determined:
– the semantics of the content and the logical structures of geographic data, this shall be done in an application schema,
– a system and platform-independent data structure that can represent data corresponding to the application schema.

The application schema is a conceptual schema for data required by one or more applications in the specified domain. The conceptual schema is a formal description of a conceptual model in the specified conceptual schema language, and the conceptual model defines the main concepts for the specific application domain.

The application schema is the basis of a successful data interchange and defines the possible content and structure of the data. It shall be accessible to both ends (sender and receiver) of a data interchange to ensure a successful result.

The application schema used for data interchange shall be written in the UML conceptual schema language, in accordance with:
– ISO/TS 19103 – defines the UML profile in the geographic information (geoinformation/geomatics) domain, adjusted to the ISO 19100 series of International Standards (established convention of naming and modelling is common for the whole series of International Standards),
– ISO 19109 – specifies the general rules for creating and documenting application schemas, including principles for the definition of features. The scope of this International Standard includes the following: conceptual modelling of features and their properties from a universe of discourse, definition of application schemas, use of the conceptual schema language for application schemas, transition from the concepts in the conceptual model to the data types in the application schema and integration of standardized.

Additionally, for the neutral spatial data interchange, the ISO 19100 series of International Standards specifies two XML (Extensible Markup Language) based encoding rules for application schemas. One of them defines ISO 19136 Annex E. It specifies a XML based encoding rule for ISO 19109 conformant application schemas that can be represented using a restricted profile of UML that allows for a conversion to GML (Geography Markup Language). GML is also a spatial data interchange format.

4. GIS Data Standard and conceptual modelling

The current version of the GIS data standard needed for the environment protection services is 3.03.01. This document includes many tables, dictionaries and symbols that should be used to present different nature forms. It also gives series of examples for previously defined objects.

Although the authors of above mentioned standard have done a tremendous work, they didn't take into account the methodology of geographic information modelling with the use of conceptual schemas that are the base for the coherent GIS realizations on different software-hardware platforms, thereby enabling GIS interoperability (Pachelski 2003).

In discussed GIS data standard, there were defined various tables that may suggest a relational implementation environment for proposed standard. Unfortunately such standard notation denies the idea of interoperability and can make its implementation difficult for different software-hardware platforms in the future.

The GIS data standard contains 12 tables describing points that differ form each other only names, e.g. "National parks boundaries points" or "Nature reserves boundaries points" (Fig. 2 and Fig. 3) and hence suggests a few remarks.

Punkty załamania granic parków narodowych parknaro_pft							Geometria _Punkt_
Nazwa pola	**Typ danych**	**Dozw. PUSTE**	**Domena / Format**	**Precyzja**	**Skala**	**Długość**	**Opis**
guid_	GUID	Nie					Identyfikator unikatowy globalnie **GUID** (np. {530F3B5C-0156-49F4-B536-89AC1BE75C8B}) formy ochrony przyrody (identyfikator musi mieć jednakową wartość w klasie obiektów parknaro_pft, parknaro_lft, park naro_aft, otulparknaro_pft, otulpark naro_lft, otulparknaro_aft dla tej samej formy ochrony przyrody)
nr	Short	Nie		5			**Numer** punktu załamania
x	Double	Nie		10	4		**Rzędna X** w układzie współrzędnych płaskich prostokątnych „1992" wyrażona liczbą zmiennopozycyjną (np. 378088,5872)
y	Double	Nie		10	4		**Odcięta Y** w układzie współrzędnych płaskich prostokątnych „1992" wyrażona liczbą zmiennopozycyjną (np. 519044,8505)
Uwaga: Dane powinny zostać automatycznie wygenerowanie z wektora granicy parku narodowego							

Fig. 2. "National parks boundaries points" table; source: Łochyński, Guzik 2009

To fill these tables, there will be needed data delivering, definitely in most cases, by land surveyors. That's why maybe it would be a good idea to adopt data types describing points, boundaries and areas (in geometric part), e.g. "OMG_Punkt", "OMG_Granica" or "OMG_Linia"[1] from geodesists. Also an assumption concerning the number of fractional digits seems to be wrong. What kind of points will have coordinates specifies with 4 fractional digits? The ISO/TS 19103 recommends the "Real" type (ISO/TS 19103:2005) for real values and the number of fractional digit should conform the point coordinates accuracy.

---

[1] The abbreviation OMG stands for "OgólnyModelGeodezyjny" – general geodesy model (Pachelski, Parzyński 2007).

Therefore it is only possible to specify the acceptable point coordinates accuracy or the acceptable point position accuracy, but the standard authors didn't design such column in tables.

The next contentious issue is an identifier. According to the INSPIRE directive in the spatial data infrastructure each object must have its own unique identifier. However in the description of globally unique identifier, the authors of discussed standard impose the same value for the unique identifier for several object classes (Fig. 2 and Fig. 3).

Punkty załamania granic rezerwatów przyrody rezeprzy_pft							Geometria *Punkt*
Nazwa pola	**Typ danych**	**Dozw. PUSTE**	**Domena / Format**	**Precyzja**	**Skala**	**Długość**	**Opis**
guid_	GUID	Nie					Identyfikator unikatowy globalnie **GUID** (np. {530F3B5C-0156-49F4-B536-89AC1BE75C8B}) formy ochrony przyrody (identyfikator musi mieć jednakową wartość w klasie obiektów rezeprzy_pft, rezeprzy_lft, rezeprzy_aft, otulrezeprzy_pft, otulrezeprzy_lft, otulrezeprzy_aft dla tej samej formy ochrony przyrody)
nr	Short	Nie		5			**Numer** punktu załamania
x	Double	Nie		10	4		**Rzędna X** w układzie współrzędnych płaskich prostokątnych „1992" wyrażona liczbą zmiennopozycyjną (np. 378088,5872)
y	Double	Nie		10	4		**Odcięta Y** w układzie współrzędnych płaskich prostokątnych „1992" wyrażona liczbą zmiennopozycyjną (np. 519044,8505)
Uwaga: Dane powinny zostać automatycznie wygenerowanie z wektora granicy rezerwatu przyrody							

Fig. 3. "Nature reserves boundaries points" table; source: Łochyński, Guzik 2009

Further analysing proposed GIS data standard in the environment protection domain, one of columns in table concerning the point coordinates accuracy is called "Scale". This concept has quite different meaning in geodesy and this column just relates to geodetic data.

In case of various objects boundaries (Fig. 4), there is the similar situation as described above for points. The GIS data standard also contains 12 tables describing boundaries that differ form each other only names. For this reason the foregoing remarks also refer to tables describing boundaries.

Moreover there is no relation between boundaries and points that define them. In the current version of discussed standard the role of such relation plays an identifier that in standard authors' opinion should be the same for a few object classes describing the same nature form. In that case how to specify that one point will belong to several different boundaries? Will it have several different identifiers or there will be a few points with the same coordinates but different identifiers in a database? Above described situation leads to data redundancy, what is an undesirable property (and even unacceptable) of the relational database.

Granice parków narodowych							Geometria	*Polilinia*
parknaro_lft								

Nazwa pola	Typ danych	Dozw. PUSTE	Domena / Format	Precyzja	Skala	Długość	Opis	
guid_	GUID	Nie					Identyfikator unikatowy globalnie **GUID** (np. {530F3B5C-0156-49F4-B536-89AC1BE75C8B}) formy ochrony przyrody (identyfikator musi mieć jednakową wartość w klasie obiektów parknaro_pft, parknaro_lft, park naro_aft, otulparknaro_pft, otulpark naro_lft, otulparknaro_aft dla tej samej formy ochrony przyrody)	
zrdopsgrn	Short	Nie	zrdopsgrn	1			**Źródło opisu granic**	
Uwaga: –								

Fig. 4. "National parks boundaries" table; source: Łochyński, Guzik 2009

The discussed standard also contains many tables describing different types of areas. A few attributes occurs in all tables, some of them in most tables, but not in all of them. This is a perfect occasion to use the UML generalization relationship. Above mentioned remarks can also refer to tables concerning spatial arrangement of flora and fauna species. Tables: "Species standings", "Species standings observations", "Standings and occurrence areas observations" need to apply suitable UML relationships, in this generalizations or simple associations that give the opportunity to indicate relations between objects like area or boundary and given flora and fauna species.

5. GIS Data Standard according to the ISO 19100 series of International Standards

The adaptation of the developed GIS data standard for the environment protection services to the requirements of the ISO 19100 series of International Standards needs first of all applying the methodology of geographic (spatial) information modelling with the use of conceptual schemas. First, for the application domain, in this case environment protection, there should be worked out the general conceptual model in UML, so called application schema. Next, this model can be transformed into the logical model that conforms to requirements of the chosen software-hardware platform. The last stage is the physical model creation, therefore an empty database structure generation in the specified software-hardware environment (Fig. 5).
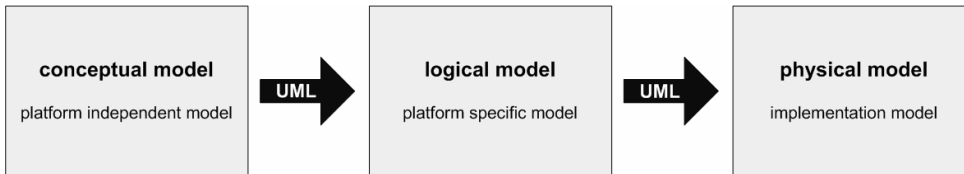
Fig. 5. Stages of system modelling; source: own elaboration

The use of formal mean of geographic information modelling (UML) in the process of developing this GIS data standard would permit for agile dedicated GIS designing and creating according to this standard in the future. Compared to the discussed standard specification presented by its authors, UML allow to much more "compact" and readable record of the general idea of the GIS data standard. Instead of dozens pages with tables, the same can be showed with the aid of one simple UML package diagram (Fig. 6) or several UML class diagrams (Fig. 7).

During UML class diagrams designing the best method is to use CASE (Compuetr Aided Software Engineering) tools that facilitate the process of following system models building. These tools (e.g. Sparx Systems Enterprise Architect) also let directly use the recommendations of the ISO 19100 series of International Standards or the INSPIRE directive. When working on the specified application schema it is possible to use the normalized models (coming form the ISO standards) or INSPIRE data specifications by loading some special repositories with such models in CASE tool (Fig. 8).
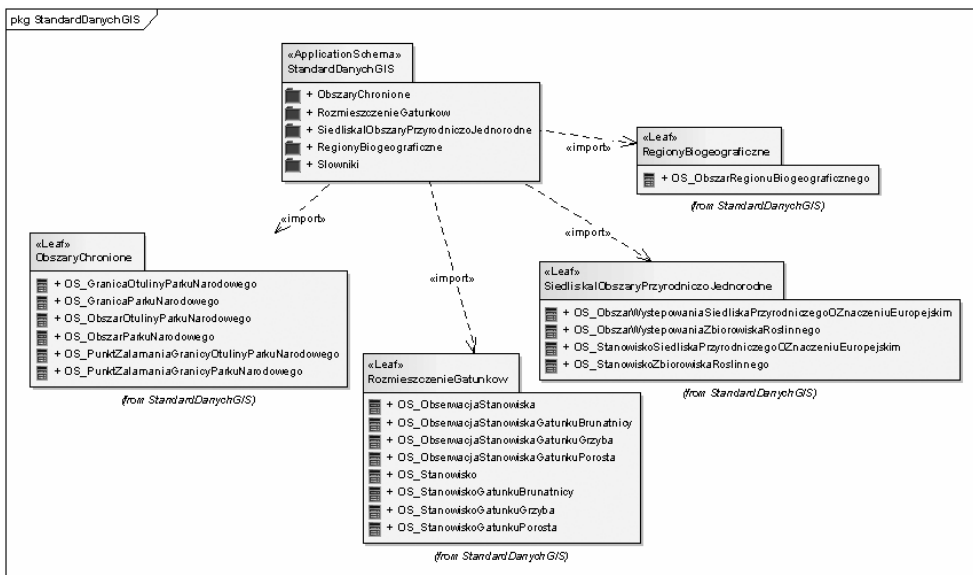

Fig. 6. The GIS data standard as an UML package diagram (fragment); source: own elaboration
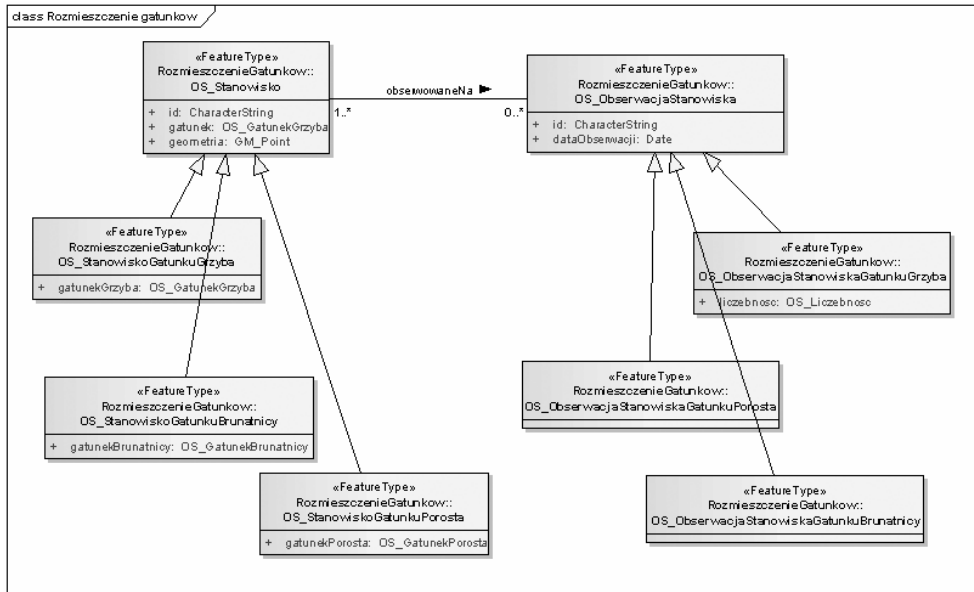
Fig. 7. The GIS data standard as an UML class diagram (fragment); source: own elaboration
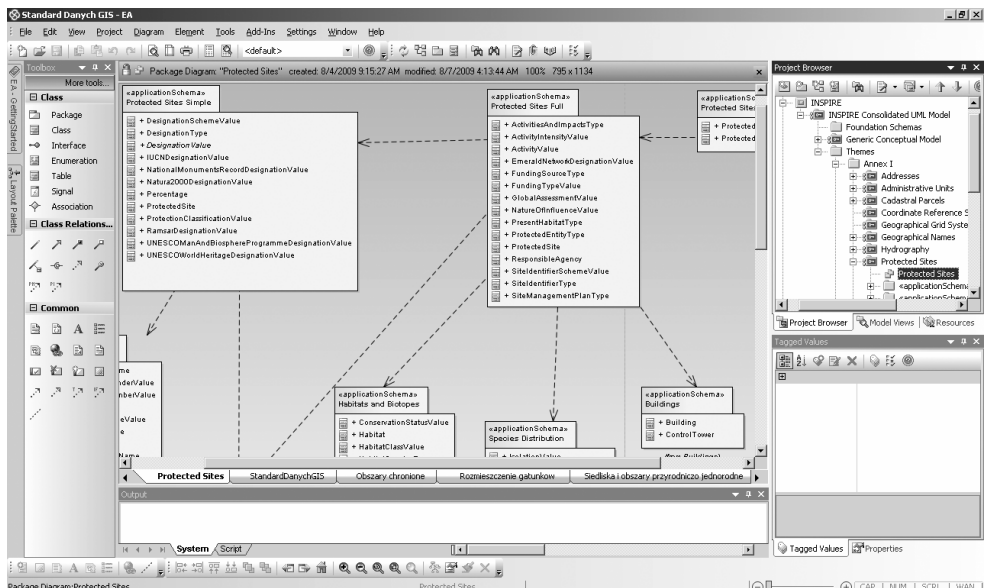

Fig. 8. The INSPIRE models repositories in the Enterprise Architect software; source: own elaboration

The another benefit of applying UML to develop the GIS data standard for the environment protection is the possibility of writing it in GML (Fig. 9) that is also an established open interchange format for spatial data, according to the GIS interoperability principle.

```xml
<?xml version="1.0" encoding="UTF-8" ?>
- <xs:schema xmlns:os="StandardDanychGIS4OchronaSrodowiska" xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns:gml="http://www.opengis.net/gml/3.2" targetNamespace="StandardDanychGIS4OchronaSrodowiska"
    elementFormDefault="qualified" attributeFormDefault="unqualified">
    <xs:import namespace="http://www.opengis.net/gml/3.2"
      schemaLocation="http://schemas.opengis.net/gml/3.2.1/gml.xsd" />
    <xs:element name="OS_ObszarOtulinyParkuNarodowego" type="os:OS_ObszarOtulinyParkuNarodowegoType"
      substitutionGroup="gml:AbstractFeature" />
  - <xs:complexType name="OS_ObszarOtulinyParkuNarodowegoType">
    - <xs:complexContent>
      - <xs:extension base="gml:AbstractFeatureType">
        - <xs:sequence>
            <xs:element name="id" type="xs:string" />
            <xs:element name="powierzchniaFormyOchronyPrzyrody" type="xs:double" />
            <xs:element name="powierzchniaMatematyczna" type="xs:double" />
            <xs:element name="geometria" type="gml:SurfaceType" />
          - <xs:element name="obszar" type="os:OS_ObszarParkuNarodowegoPropertyType" maxOccurs="unbounded">
            - <xs:annotation>
              - <xs:appinfo>
                  <gml:reversePropertyName>os:otulina</gml:reversePropertyName>
                </xs:appinfo>
              </xs:annotation>
            </xs:element>
          </xs:sequence>
        </xs:extension>
      </xs:complexContent>
    </xs:complexType>
  - <xs:complexType name="OS_ObszarOtulinyParkuNarodowegoPropertyType">
    - <xs:sequence minOccurs="0">
        <xs:element ref="os:OS_ObszarOtulinyParkuNarodowego" />
      </xs:sequence>
      <xs:attributeGroup ref="gml:OwnershipAttributeGroup" />
      <xs:attributeGroup ref="gml:AssociationAttributeGroup" />
    </xs:complexType>
```

Fig. 9. The GIS data standard in the GML format (fragment); source: own elaboration

It is also worth to notice that the discussed GIS data standard was intended to implement in the ESRI ArcGIS environment. And this is the next argument for applying UML to this standard elaboration because there exists a simple "technological path" that enables the automatic transition UML-ArcGIS (Chojka 2009).

ESRI is the only GIS software producer that makes possible to use directly the UML to design GIS database (Gajc 2004). To this end the ArcInfo UML model template for CASE tools is used: Microsoft Visio or Rational Rose (since 2003 Rational solution is not supported by ESRI).

Therefore the only thing to do is to adjust the GIS data standard conceptual model, prepared previously in UML, to geodatabase data model in the ArcGIS system by applying the ArcInfo UML model template for the Microsoft Office Visio Professional 2003 (Chojka 2008). This template contains a set of classes that enable to locate designed data model in the object structure of ArcInfo data model (Fig. 10). Then, ready data model (ArcGIS conforming logical model) designed in UML is exported to XMI format (XML Metadata Interchange). The proper ArcCatalog tool generates the database structure on the ArcGIS platform form the XMI file (Chojka 2008).
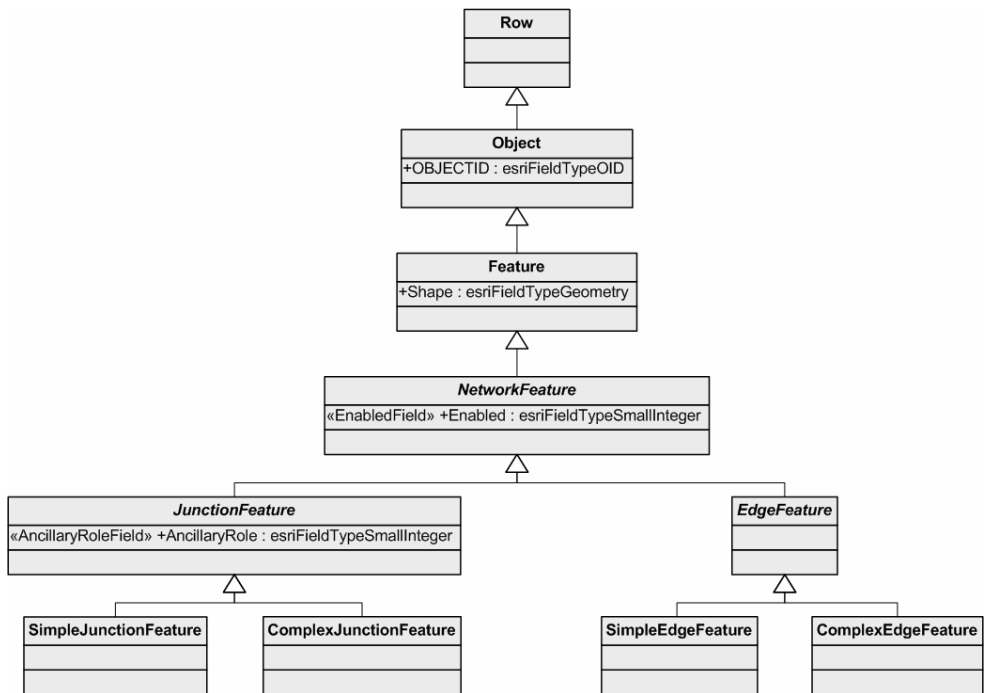
Fig. 10. The basic classes of the ArcInfo object data model; source: own elaboration.

6. Summary

The GIS data standard in the environment protection domain to really could become a standard, first of all ought to take into consideration the recommendations of the ISO 19100 series of International Standards in the geographic information (geoinformation/geomatics) domain, so use UML – the formal mean of describing the geographic (spatial) information. However except this notation, very significant is the methodology of geographic information modelling with the use of conceptual schemas that are the base for the coherent GIS realizations on different software-hardware platforms, thereby enabling GIS interoperability.

The adaptation of the discussed GIS data standard for the environment protection services to the requirements of the ISO 19100 series of International Standards can bring a lot of benefits, because UML will:
  – ensure much more "compact" record of the GIS data standard idea,
  – let first of all order the elementary concepts from the specified application domain (environment protection),
  – enable to quick transform the ready standard into the GML format,
  – let fast standard implementation in the ESRI ArcGIS environment,

– make possible to use some appropriate tools that let directly use both the recommendations of the ISO 19100 series of International Standards and the INSPIRE directive.

References

Chojka A., 2008. Projekt podsystemu GIS dla gminy i rzeczoznawcy majątkowego. Rozprawa doktorska. Uniwersytet Warmińsko-Mazurski w Olsztynie, Wydział Geodezji i Gospodarki Przestrzennej.

Chojka A., 2009. Adaptacja technologii MDA do budowy systemu geoinformacyjnego na poziomie gminy. Roczniki Geomatyki, T. VII, Zeszyt 5 (35), ss. 25-39, Warszawa.

Directive, 2007. 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). Official Journal of the European Union (25.04.2007).

Gajc B., 2004. UML w akcji. O praktycznym zastosowaniu schematów pojęciowych. Geodeta, Magazyn Geoinformacyjny nr 10 (113) ss.36-37.

ISO/TS 19103:2005 Geographic information – Conceptual schema language

ISO 19109:2009 Geographic information – Rules for application schema

ISO 19136:2007 Geographic information – Geography Markup Language (GML)

Łochyński M., Guzik M., 2009. Standard danych GIS w ochronie przyrody. Wersja 3.03.01. Poznań-Zakopane-Kraków.

Protected Sites, 2010. INSPIRE Data Specification on Protected Sites – Guidelines, 2010. http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.1.pdf

Pachelski W., 2003. Normalizacja w dziedzinie informacji geograficznej. Normalizacja, Nr 11, ss. 18-26.

Pachelski W., Parzyński Z., 2007. Aspekty metodyczne wykorzystania norm serii ISO 19100 do budowy georeferencyjnych składników krajowej infrastruktury danych przestrzennych. Roczniki Geomatyki, t.V, z.3, Warszawa.

prCEN/TR 15449, 2006. Geographic information – Standards, specifications, technical reports and guidelines, required to implement Spatial Data Infrastructure.

GIS Standard, 2010. Standard danych GIS w ochronie przyrody – wspólny język służb ochrony przyrody w gromadzeniu i wymianie informacji geograficznej. Materiały Konferencji „Geoinformacja w zarządzaniu środowiskiem", 25.05.2010, Warszawa.

Agnieszka Chojka

University of Warmia and Mazury in Olsztyn
Faculty of Geodesy and Land Management
Chair of Land Surveying and Geomatics