Conceptual Modelling of Real Estates for the Purposes of Mass Appraisal*

Ewa DĘBIŃSKA and Piotr CICHOCIŃSKI, Poland

Key words: valuation, Unified Modelling Language (UML), CASE, geodatabase, conceptual modelling

SUMMARY

Legislative works are taking place in Poland on elaboration of detailed regulations concerning mass appraisal. The approval of them will cause necessity of valuation of several million properties.
The basic sources of information pertaining to real estate valuation include: the real property register and the land and buildings cadastre which are located in different institutions. As a result, gathering all the documentation required for a real estate valuation becomes a time-consuming task.
In the paper the authors propose a database model for real estate valuation. It contains information regarding types of real estate such as parcels, buildings and dwelling areas. We currently live in a computerized society, therefore the planned database must not only consist of descriptive attributes, but also contain information regarding their locations and shapes. Such a database, containing geographical information, that distinguishes it from among other databases, can be named a geodatabase. The introduction of the geodatabase model allows the application of the methods of designing relational databases, including CASE tools, also for geographic information.
The key stage of database design is conceptual modelling. During conceptual modelling entities and their properties that are important for mass appraisal purposes and their relationships were defined and the behaviours that apply to these entities were identified. Such conceptual model may exist only in the minds of people and be communicated verbally and often imprecisely. But it may also be written down using conceptual schema language and stored for wider dissemination. A conceptual schema language provides the semantic and syntactic elements used to describe the conceptual model rigorously in order to convey meaning consistently. Because a conceptual schema language provides a uniform method and format for describing information, it is possible to read and update the resulting conceptual schema by computer systems as well as human beings. The conceptual schema language used for developing this model was UML. The UML is a language for specifying, visualizing, constructing, and documenting the artefacts of software systems, as well as for business modelling and other non-software systems. Application of Microsoft Visio software for "drawing" of UML schema allowed the use of CASE Tools from ESRI to convert it into logical model and subsequently to automatically generate in ArcGIS software appropriate database structure, ready for data input.

* This work is financed from funds for science allocated for year 2007 at AGH University of Science and Technology

Conceptual Modelling of Real Estates for the Purposes of Mass Appraisal†

Ewa DĘBIŃSKA and Piotr CICHOCIŃSKI, Poland

1. INTRODUCTION

Legislative works are taking place in Poland on elaboration of detailed regulations concerning mass appraisal (Telega et al., 2002). The approval of them will cause necessity of valuation of several million properties. Diversity of the property markets and their still emerging character does not always allow the application of advanced valuation methods, as well as direct implementation of valuation methods used in countries with the fully developed market economy. One of the basic methods of real estates market value determination is the valuation accomplished by the means of comparative approach. This approach makes assumption, that the market value of real estate is assessed through the comparison to other similar real estates, for which transaction prices and characteristics distinguishing these real estates and having the essential influence on their value are known.

The main information sources used for the valuation of real estates are: land and buildings cadastre, land and mortgage registers, local spatial development plans, basic maps, spatial registration of utility infrastructure, registers of building permits, builder's records and resources of the Central Statistical Office.

"The following are characteristic attributes of grounds built-up or marked out for development, and grounds marked up for other purposes than farming and forestry, that influence the cadastral value:

- location;
- intended purpose specified in the local spatial development plan or, in the case of lack of this plan, the manner of using;
- level of outfitting with technical infrastructure equipment;
- state of development;
- the soil-based land classification, if specified in the land and buildings cadastre" (Rozporządzenie, 2005).

Theoreticians and practitioners in the area of the real estates valuation (Czaja, Parzych, 1999; Czerkies, 1998) indicate a few essential attributes yet:

- area,
- location (the zone),
- geometrical conditions (the shape, the form of the ground),
- surroundings (neighbourhood),
- communication accessibility,
- popularity of given location.

---

† This work is financed from funds for science allocated for year 2007 at AGH University of Science and Technology

It is assumed that the majority of data required for valuation will come from the mentioned above sources. At first sight it seems that such situation really most often will take place: information on location, the shape, the soil-based land classification, the function and characteristics of building should be recorded in the land and buildings cadastre; destination of the ground is described in the local spatial development plan; the accessibility of utilities is described by the spatial registration of utility infrastructure.

But unfortunately the access to certain data not necessarily gives the possibility of their immediate use, which means that in the most cases that data itself does not constitute attributes of real estates. Some attribute values result not from properties of objects, but from relationships between them. Other ones are the derivatives of certain characteristics (for example geometrical) of real estates. These indirect real estates attributes can be determined with the use of analytic functions of GIS software (Cichociński, Parzych, 2006), but realization of this operation demands necessary data to be stored in one database.
The main objective of this paper is presentation of database design process along with final database model. It should contain information regarding types of real estate such as parcels, buildings and dwelling areas. We currently live in a computerized society, therefore the planned database must consist not only of descriptive attributes, but also contain information regarding real estates location and shape.

2. GEODATABASES

Until quite lately the storage of such information along with descriptive data in the coherent database was difficult. These two types of information were divided into two separate files joined with unique identifiers. For example in ArcInfo coverages spatial data were recorded in indexed binary files, while attributes were stored in relational database called INFO. Nowadays convenient solution for databases containing geographical information seems to be object-relational data model.

The authors use in this paper the term geodatabase for description of databases storing the geographical information with the utilization of object-relational data model. This name was originally used by ESRI to describe their own data model based on object-relational database, however in authors opinion it adequately reflects the general idea of using object-relational databases for recording of the complete (geometrical and descriptive) information about geographical objects. In this paper, as example, the ESRI solution is described, because it is supported by series of tools automating the transition process from conceptual model to physical implementation.

Geographical data are complex. The shapes of line and area objects are structured sets of coordinates that do not record themselves properly in fields of standard types, such as: integer, real or string, so they require the suitable model of the database. Nowadays three database models dominate on the market: relational, object-oriented and object-relational. In the relational model every object and its attributes are recorded as row in the table, and the information in the single cell is atomic. However despite its clear rules, it is not suitable for geographical data. In the object-oriented model data are stored in the form of objects which can be read only by methods made available by these objects. Object-relational model bridges the gap between relational and the object-oriented models of the database. Data are still stored in tables, but attributes can have values of non-standard types. This is why in case of geographical data the best solution seems to be object-relational database model extended to the geodatabase model.

Geodatabase, keeping the information in the object-relational database, allows the data to be manipulated as the set of objects, simultaneously using relational database as the internal method of data storage. Practically geodatabase is a physical representation of objects existing in the real world. It makes possible the storage of spatial data (geometrical, descriptive, gridded) in the database management system (DBMS). An advantage of geodatabase, as the format of data storage, is the free access which allows users to create, utilize and manipulate geographical data. Thanks to the open technology and the publicly accessible documentation the use of geodatabase does not require software from particular producer. Additionally data modelling in geodatabase with the utilization of the object-oriented methodology (Zeiler, 1999) makes possible description of reality in more natural way. Such approach provides better understanding of objects mutual influences.

The introduction of the geodatabase model allows the application of relational databases design methods also for geographic information. Three stages of database design can be distinguished (Connolly, Begg, 2004): the construction of conceptual data model consisting of acquisition of information about objects, relationships and attributes, then the creation of logical model of a database, which means the transformation of a conceptual data model into logical database structure, and at last physical implementation of a logical data model. Models: conceptual and logical refers to the users view, and the model of database implements the data model within the framework of the relational database technology.

3. CONCEPTUAL MODELLING

Conceptual modelling is the key stage of database design. Creation of conceptual model encompasses variety of intellectual processes and images referring to the project. It is necessary to precisely imagine the problem and the method to solve it.

A key task in the process conceptual model building is the precise definition of objects of interest and identification of relationships among them. As examples of objects streets, lots, owners and buildings can be given. Examples of relationships among them can be “located on”, “owned by”, “is a part of”.

The proper selection of objects on the stage of design is not only essential for preserving the proper (informative) function of planned database. It has also influence on the efficiency of data administration and particularly on execution of database queries and spatial analyses (Makowski, 2005).

Figure 1 presents relationships between the real world and conceptual schema. Defining of concepts from the universe of discourse, being the part of the real world, takes place in the conceptual model. These mentioned concepts are delivered by conceptual formalism, which can be treated as set concepts from area of modelling, used to describe the conceptual model.

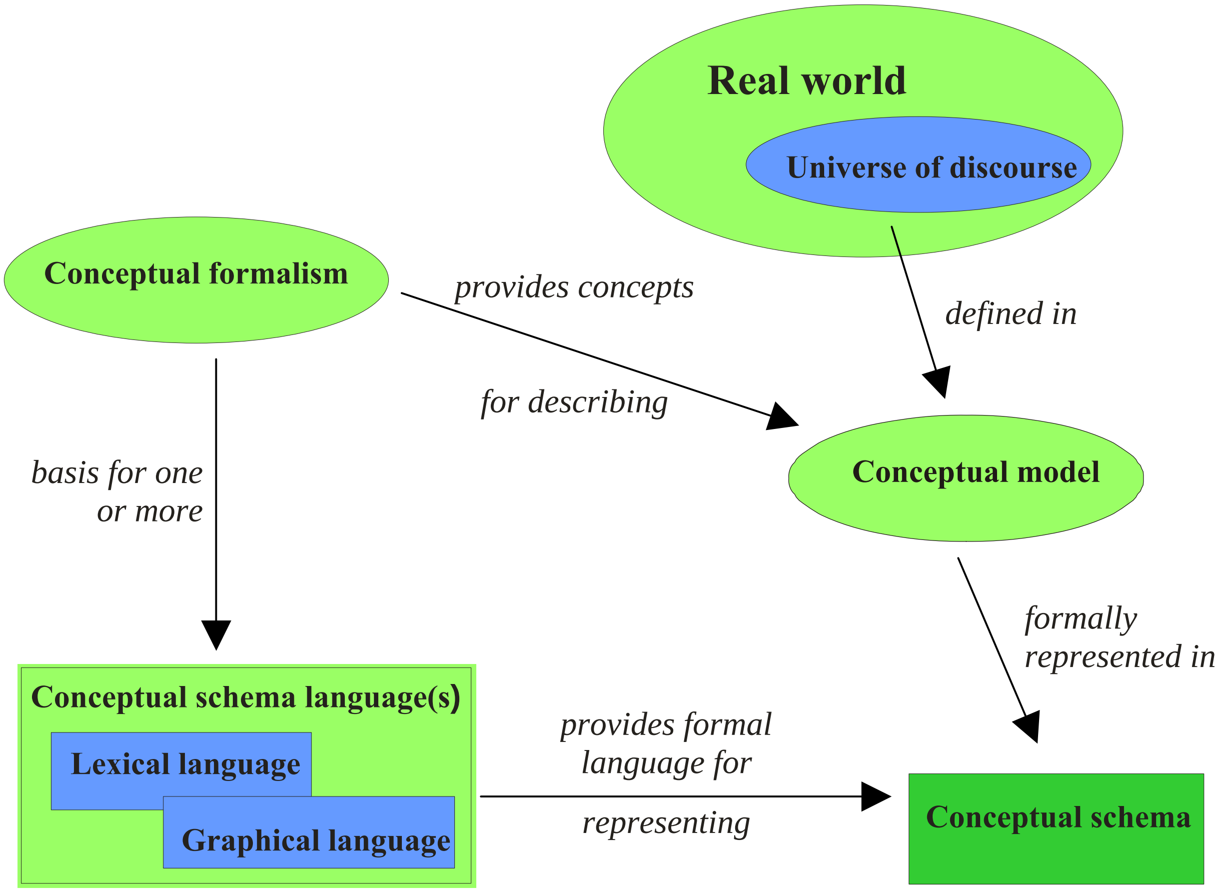

Fig. 1. From the real world to conceptual schema (ISO 19101:2002).

Such conceptual model may exist only in the minds of people and be communicated verbally and often imprecisely. But it may also be written down using conceptual schema language and stored for wider dissemination. A conceptual schema language provides the semantic and syntactic elements used to describe the conceptual model rigorously in order to convey meaning consistently. Because a conceptual schema language provides a uniform method and format for describing information, it is possible to read and update the resulting conceptual schema by computer systems as well as human beings. UML is the most popular conceptual schema language nowadays.

4. UNIFIED MODELLING LANGUAGE

The Unified Modelling Language (UML) is a language for specifying, visualizing, constructing, and documenting the elements of software systems, as well as for business modelling and other non-software systems. The idea behind UML is to make possible the use of simple means for creation of graphic model presenting any fragment of the reality. This is why it is very good tool for accurate expression of thoughts in the form of models. Simplifying, the UML notation is the set of readable symbols and signatures, which can be understood even by persons with the minimum knowledge of computer science.

In 1996 Object Management Group (OMG) approved UML as the standard language for object-oriented methods. This fact had a significant influence on increase of the UML popularity among programmers. UML was also adopted as a formal language for description of models and conceptual schemata in the series of ISO 19100 standards referring to geographical information (ISO 19101:2002).

Building a database schemata using UML requires the knowledge of several basic definitions from the domain of object-oriented design:
The object – has many meanings both in computer science and in the real world. The most accurate seems to be the description which defines the object as a concrete entity which can be univocally identified and compared to the real or material entity. On the other hand, the object constitutes certain programmatic abstraction created for the purpose of desirable logic. The object is a basic concept, from which originate the following concepts:
The class – it states the group of objects which have identical set of attributes, operations and methods, and also relates to other groups of objects (Muller, 2000; Śmiałek, 2005).
Inheritance, generalization – the relationship between object classes describing the transfer of characteristics (attribute definitions, methods) from the superclass to its subclasses. It is a basic mechanism allowing the repeated use (Subieta, 1999).
The association – the kind of relationship between classes projecting the existing relationship between appropriate entities in analyzed objective domain (Śmiałek, 2005).
The aggregation – the relationship between object classes modelling proportion of the whole to its part. A relationship of aggregation exists between objects, if one of them can be considered as the part of the other (Subieta, 1999).

5. CONCEPTUAL MODEL OF DATABASE FOR MASS APPRAISAL

During conceptual modelling entities and their properties that are important for mass appraisal purposes and their relationships were defined and the behaviours that apply to these entities were identified. In case of data model for mass appraisal the main objects will be buildings, dwellings, parcels, in other words objects which will be subjects of valuation. There are also important sets of objects essential for the process of appraisal, having the influence on the cadastral value of real estates. This second group should contain:

- spatial development plan,
- soil-based land classification,
- land use,
- shape of the terrain (form of surface),
- utility networks,
- road network,
- nearby objects, both increasing and decreasing real estate value.

A basic object in the model of database for real estates valuation is parcel understood as land real estate. The land real estates are defined in item 46 of the civil code (Ustawa, 1964): „Real estates are parts of the earth surface constituting the separate subject of ownership (grounds)". On the contrary item 48 describes buildings and other appliances permanently connected to the ground as components of land real estates (Ustawa, 1964). The conclusion from above rules is that the building should be treated in the model as the attribute of parcel. However in such solution the location of the building is skipped and the existence possibility of building real estate is not taken into account (the case when the building constitutes subject of ownership separate from the ground). Therefore the best solution seems to be treatment of the building as the separate geographical object. The one-to-many relationship should be established, allowing zero, one or more buildings to be located on a parcel.

Somehow similar problem refers to dwelling. The only difference is that dwelling does not have location, but also should be separate object (related to building) because of the possibility of dwelling real estate existence.

Spatial development plans, land use, soil-based land classification give information about real estate belonging to areas of some superior partition of geographical space. The relationship should be created describing the assignment of given real estates to particular zone. Additional detailed considerations are required in case when the given property is not entirely contained in the element of given partition. The selection of zone in which resides greater part of real estate may be one of the solutions. Another one can be establishment of relationship many-to-many, with attributes containing information about percentage of parcel area belonging to given zone.

Because Digital Terrain Models (DTMs) can not be stored in object-relational databases at this moment, it is necessary to use another method of terrain description. It can be for instance use of triangles with two attributes: the slope and the aspect – the direction of steepest downhill slope. Similarly to the other objects of the zonal character it would be necessary to tie the parcel to the triangle, wherein its greater part resides, as the main determinant of its shape of the terrain.

The road network is necessary for determination of the communication accessibility of real estates (Cichociński, 2006). According to regulations about real estates management (Ustawa, 1997) every parcel must have provided driveway. Thereby it would be advisable to relate parcel to the road or street it is located along. It can easily be done using address information. The database should be created gathering information about different kind of objects, which can be found in the vicinity of real estates, having the influence on their values.

It can be for instance depreciatory industrial plant or appreciating park. The distance and azimuth from parcel to these objects (stored as many-to-many relationship attributes) will allow estimation of their influence on real estate value. The information about parcel vicinity, existing or planned, can be obtained from the local spatial development plan.

Preceding considerations led the authors to the model depicted in the following figure.

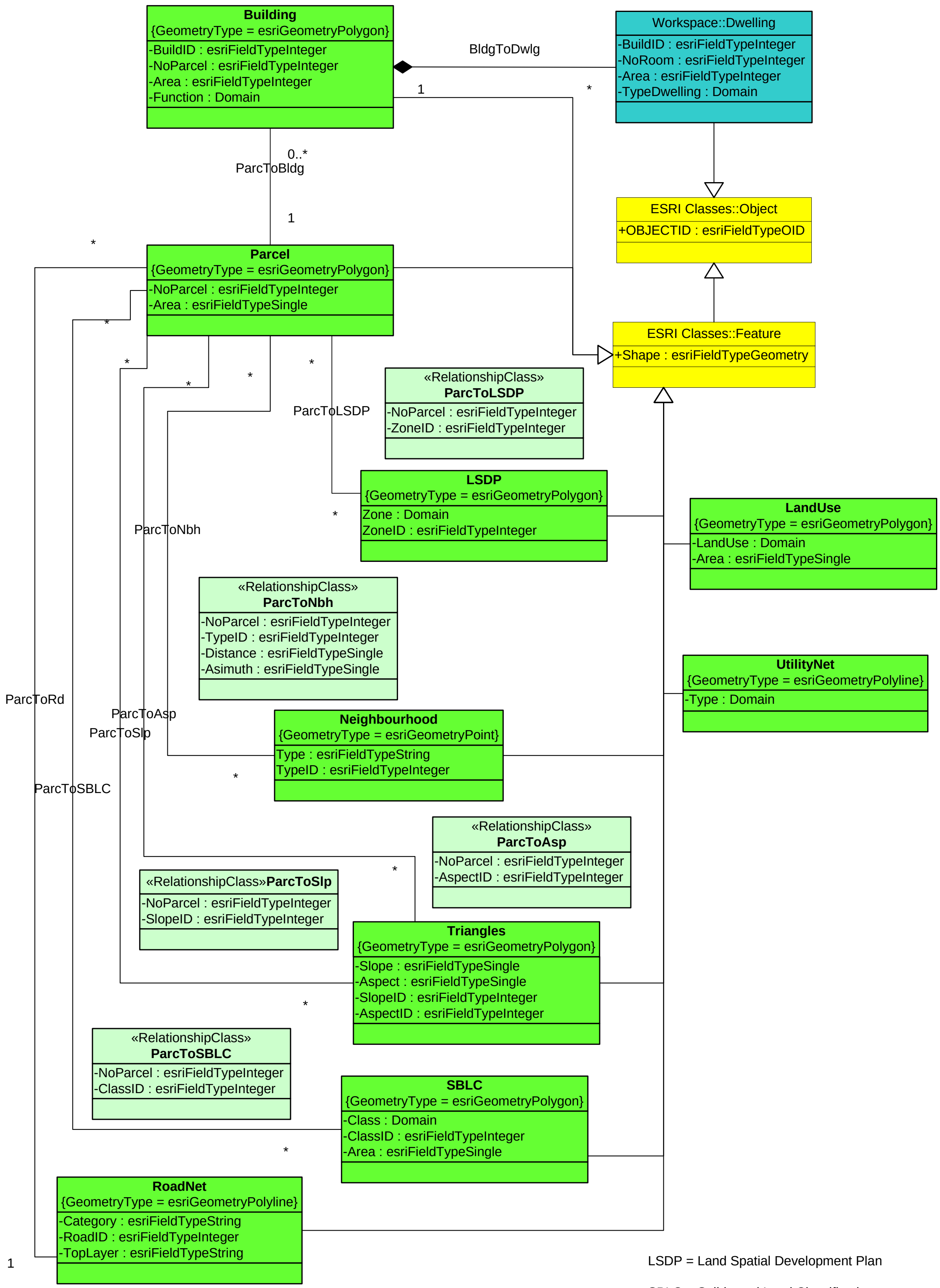

Fig. 2. Conceptual model of database for mass appraisal.

6. FROM CONCEPTUAL MODEL TO GEODATABASE

Of course the creation of conceptual model is not the last step of database design. The next one is logical database design consisting of the conceptual model transformation into the logical model with the regard to the data model in the destination database (for instance the relational model). On this stage the registration method of the descriptive attributes, spatial properties of objects and the relationships between them are defined. The logical data model is an abstraction of the objects occurring in a particular application. This abstraction is converted into database elements. An object represents an entity such as a building, a parcel or an owner and is recorded as a row. An object possesses a set of attributes, which characterize its properties, such as name, a measure, a classification or an identifier (a key) to another object. Attributes are recorded in a database in columns (called also fields). A set of similar objects forms a class, stored in a database as a table. In the past logical data models were often drawn in form of so called entity-relationship diagrams. Lately however, Unified Modelling Language (UML) achieved the greatest popularity also here.

From the logical data model a physical database model is built. Previously, most often a specialist in relational databases received the logical data model and used database administration tools to define the database schema and create new database ready for data transfer and entry. But now users have at command one more possibility: the CASE (Computer Aided Software Engineering) tools. The logical model of database recorded using UML, can be used to automatically generate the schema of database conforming to the particular specification. In the following example Microsoft Office Visio 2003 was used to record UML schema. When ready, this schema served as a basis for automatic generation of database structure in ArcGIS software form ESRI (Perencsik et al. 2004a).

The creation of the schema consisted of placing on the drawing page pre-drawn shapes from stencils. Visio delivers eight stencils. This number is closely related to the UML specification, where eight basic diagrams are defined (Use-Case, Static, Sequence, Collaboration, Statechart, Activity, Component, and Deployment diagrams). In the static structure diagram, which is intended for database design, suitable object classes were defined. Subsequently, they were given attributes with appropriate variable types. From the ArcInfo data model, made available by ESRI, object classes inherit the following attributes: ObjectID – the individual identifier of every object and Shape – containing the geometry. Additionally, relationships between classes were defined.

Prepared database schema was exported to XML file using enclosed ESRI XMI Export macro (Gajc 2004). Correctness of created XML file was checked by Schematic Checker macro. The last step was import of the correct XML file to the geodatabase using ArcCatalog Schema Wizard. This process was documented by automatically generated report. As a result empty geodatabase was created, having structure conforming to the project, ready for data input.

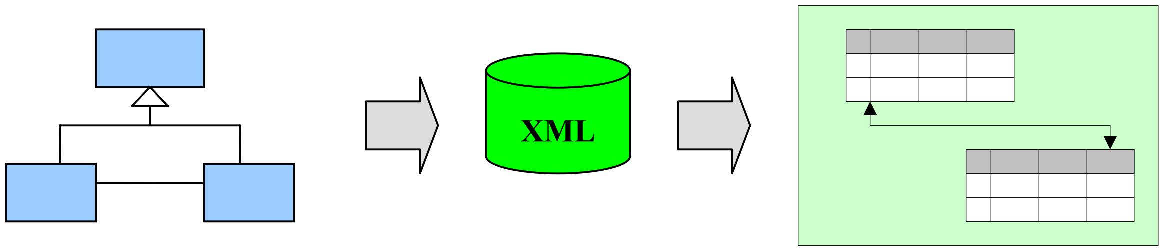

Fig. 3. Schema of UML model conversion into the geodatabase (Perencsik et al. 2004b)

7. CONCLUSION

Mechanisms presented in this paper allow for fast and efficient construction of the conceptual model for any problem. But more important thing is that tools exist making possible recording of such model by meanses of computer science, doing any modifications and finally automatically generating appropriate database structure, ready for data input. Therefore the database creation process, which was shown on example related to preparation of the mass appraisal in Poland, is straightforward. There is however extremely critical point: the first stage – conceptual modelling – has to be carried out correctly, with a regard to all objects, their attributes and relationships important for the given problem. And it has no concern with accessible methods and tools, but rather with the deep knowledge and the experience of the designer in the given field.

REFERENCES

Cichociński P. 2006, Modelowanie dostępności komunikacyjnej nieruchomości jako atrybutu niezbędnego w procesie wyceny (Modelling of real estates communication accessibility as an essential attribute in the valuation process). Roczniki Geomatyki, Tom IV, Zeszyt 3, Warszawa.

Cichociński P., Parzych P. 2006, Application of Geographic Information Systems for Real Estate Valuation Support, XXIII FIG Congress, Munich, Germany, October 8–13, 2006

Connolly T., Begg C. 2004, Systemy baz danych. Praktyczne metody projektowania, implementacji i zarządzania – tom 1 (Database systems. Practical methods of design, implementation and management – vol. 1). Wydawnictwo RM, Warszawa.

Czaja J., Parzych P. 1999, Kataster nieruchomości jako baza systemu powszechnej taksacji (Real estates cadastre as a basis for mass appraisal system). IX Konferencja Naukowo–Techniczna Systemy Informacji Przestrzennej. Warszawa, 14–15 września 1999 r.

Czerkies J. 1998, Propozycja organizacji podsystemu informacji na cele obsługi rynku nieruchomości i gospodarki przestrzennej (The proposal of information subsystem organization for real estates market maintenance and spatial management). VIII Konferencja Naukowo – Techniczna Systemy Informacji Przestrzennej. Warszawa, 19–21 maja 1998 r.

Gajc B. 2004, UML w akcji (UML in action). Geodeta, nr 10, Warszawa.

ISO 19101: 2002 Geographic information – Reference model.

Makowski A. 2005, System informacji topograficznej kraju (National topographic

information system). Wydawnictwo Politechniki Warszawskiej, Warszawa.
Muller R.J. 2000, Bazy danych. Język UML w modelowaniu danych (Databases. UML language in data modelling). Wydawnictwo MIKOM, Warszawa.
Perencsik A., Idolyantes E., Booth B., Andrade J. 2004a, ArcGIS 9. Designing Geodatabase With Visio. ESRI Press, Redlands.
Perencsik A., Idolyantes E., Booth B., Andrade J. 2004b, ArcGIS 9. Introduction to CASE Tools. ESRI Press, Redlands.
Rozporządzenie Rady Ministrów z 29 czerwca 2005 r. w sprawie powszechnej taksacji nieruchomości (The Order in case of real estates mass appraisal). (Dz.U.2005.131.1092).
Subieta K. 1999, Słownik terminów z zakresu obiektowości (Dictionary of terms from the area of object-orientation). Akademicka Oficyna Wydawnicza PLJ, Warszawa.
Śmiałek M. 2005, Zrozumieć UML. Metody modelowania obiektowego (Understanding UML. Methods of object-oriented modelling). Wydawnictwo Helion, Gliwice.
Ustawa z 21 sierpnia 1997 r. o gospodarce nieruchomościami (The Act – Real Estates Management). (Dz.U.1997.115.741 z późn. zm.).
Ustawa z 23 kwietnia 1964 roku Kodeks Cywilny (The Act – Civil Code). (Dz.U.64.16.93 z późn. zm.).
Telega T., Bojar Z., Adamczewski Z. 2002, Wytyczne przeprowadzania powszechnej taksacji nieruchomości (projekt) (Real estates mass appraisal guidelines (draft)). Przegląd Geodezyjny, nr 6, Warszawa.
Zeiler M. 1999, Modeling our World. The ESRI Guide to Geodatabase Design. ESRI Press, Redlands.

CONTACTS

Ms. **Ewa Dębińska**
AGH University of Science and Technology
Department of Land Information
Al. Mickiewicza 30
30-059 Kraków
POLAND
Tel. +48 12 617 23 00
Fax + 48 12 617 22 77
Email: debinska@agh.edu.pl

Dr. **Piotr Cichociński**
AGH University of Science and Technology
Department of Land Information
Al. Mickiewicza 30
30-059 Kraków
POLAND
Tel. +48 12 617 34 31
Fax + 48 12 617 22 77
Email: Piotr.Cichocinski@agh.edu.pl