

Semantic model transformation within the context of INSPIRE

Eva-Maria UNGER, Gerhard NAVRATIL, and Stefan KLOTZ, Austria

Key words: INSPIRE, Semantic model transformation, Cadastral Parcels

SUMMARY

In May 2007 the Infrastructure for Spatial Information in the European Community (INSPIRE) entered into force. This directive wants to ensure that the spatial data infrastructures of all Member States are compatible and usable in a community and cross-boundary context. Due to this directive all European Member States are now working on the harmonization of their geodata and services. The services offer the users the opportunity to find, browse, share, and download spatial data. The problem is that every state or the Federal Offices create, manage, and provide their geodata in different systems, models, and data formats. To harmonize the data models the European Commission took into force the Commission Regulation No 1088/2010 to implement transformation services for geodata until 28.12.2012. The transformation service should master a schema transformation from different source data models to the INSPIRE target data model. The study behind this paper has been dealing with 1. finding and investigating current transformation tools and their used transformation languages, 2. analyzing the source model (Cadastral Parcels in Austria), 3. analyzing the target model, and 4. transforming and validating a prototype model transformation. A general discussion about INSPIRE, the interaction between data models and reality and the basis theory about data modelling give a quick overview of the thematic. Also the principal method for model transformation and their direct use has been shown. Some problems have been identified which may occur during this process steps. The paper describes a possible prototype and gives an overview which requirements already can be fulfilled by the Austrian cadastre.

Semantic model transformation within the context of INSPIRE

Eva-Maria UNGER, Gerhard NAVRATIL, and Stefan KLOTZ, Austria

1. INTRODUCTION

The directive 2007/2/EC of the European Parliament and the Council of 14 March 2007 has established an Infrastructure for Spatial Information in the European Community (INSPIRE). The INSPIRE directive entered into force on the 15th May 2007. This directive shall ensure that the spatial data infrastructures of all Member States are compatible and usable in a community and cross-boundary context. Its main aim is to support policy-making for the protection of the environment. The European Spatial Data Infrastructure shall be based on the existing National Spatial Data Infrastructures. Due to this directive all European Member States have to harmonize their geodata and services. INSPIRE wants to achieve this harmonization through standardized data models and interoperable services. These services offer the users the opportunity to find, browse, share, and download spatial data. The problem is that every country or the Federal Offices has its own geodata in a different system, model, and format. The models may differ considerably, so that data from one state does not fit to data from another state. A schema transformation from the source model to the target model may provide the, harmonization of the data. The paper starts with a short section about the model theory. The most relevant terms are defined in this section. In section number 3 the principal methods of model transformation are described and discussed. In section number 4 the source model is analyzed. Afterwards in section number 5 the INSPIRE defined target model is analyzed.

2. Model Theory

The term model can be defined in the following way: “A model represents a system that is a part of reality.” ([Fra06] pp. 29).

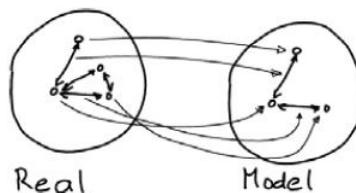


Figure 1: Interaction between reality and model [Fra06]

Therefore models can never represent reality completely. The depicted part of reality is called Universe of Discourse in a technical terminology. This part is just a detail of the real world and has been abstracted from reality. This is important for understanding models. In the beginning, this part of the model is not documented in written form. However, there is an intended goal connected with the modelling. As soon as the rules of the model are written

down, the model is defined as a conceptual model in accordance with ISO 19101 Geographic information – Reference model [ISO101]. In the next step the model is formally written down and then it is called conceptual schema. This distinction is important for the basic understanding of model theory. [Eis10]

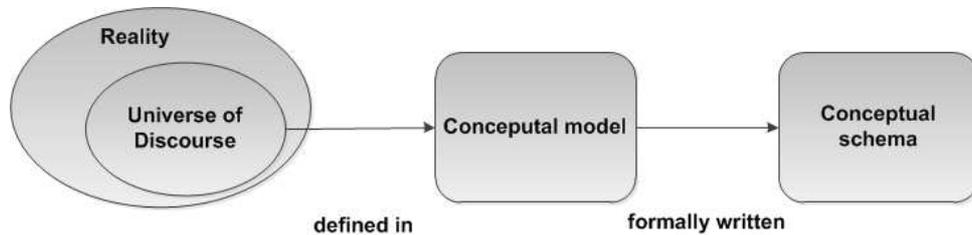


Figure 2: Transfer from Universe of Discourse to conceptual schema [Eis10]

3. Methods of model transformation

Principally there are 2 methods for model transformation of geodata: the syntactical transformation and the semantic transformation.

3.1 Syntactical transformation

Within the syntactical transformation the syntax of geodata is converted. This means that format changing is taking place e.g. shape-files are changing into GML-files. However this method of transformation is insufficient for the purpose of use. INSPIRE is not asking for format changing, in fact there is the need of transferring a local data model into a INSPIRE specified target model.

For this purpose the semantic transformation fits better. Even tough with using this method of transformation some problems may occur.

3.2 Semantic transformation

Within the semantic transformation the data is restructured, supplemented and/or reduced to fit to the target model. A semantic transformation enables exchange of heterogeneous data.

Unfortunately semantic transformation is dealing with two problems [Sta09]:

1. lossless mapping of the semantics between different data models is not possible
2. the bijectivity of mapping rules from source- to target model is not guaranteed

This problem can easily be shown with an example:

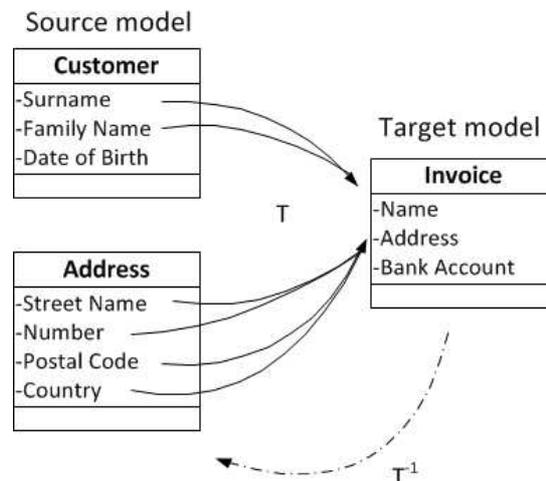


Figure 3: Problems with semantic transformation [Sta09]

The source model requires the classes "Customer" and "Address". "Customer" has the attributes: Surname, Family Name, and Date of Birth and the class "Address" consists of Street Name, Number, Postal Code, and Country. The target model is defined as the class "Invoice" with the attributes: Name, Address, and Bank Account. The attribute Name is composed of Surname and Family name. The attribute Address is composed of Street Name, Number, Postal Code, and Country. Due to the visualization in Figure 3 it is easy to identify that not all attributes of the class "Customer" can be mapped (Date of Birth) and that also not all attributes of the class "Invoice" from the target model can be filled (Bank account). Furthermore due to the transformation the semantic/structure of the data has been changed. In the source model there are 2 classes but in the target model there is just one class. Therefore information like the date of birth is lost. Additional Information, which is required in the target model, can not be filled. Due to the structural change it is not possible to retransfer the target model into the source model.

4. Source model for Cadastral Parcels in Austria

The source model for Cadastral Parcels in Austria is based on two pillars.

The first pillar is the cadastre and its technical implementation in the digital cadastral map (DKM). The DKM shows the technical aspects of the parcels. The digital cadastral map is used to provide evidence of parcel boundaries and to visualize the legal provision circumstances of the parcels. The digital cadastral map is the graphical representation of the Austrian cadastre and is using the official Austrian coordinate system. It is managed by the local surveying departments and shows [Aba11, pp. 141ff]

- the borders of parcels,
- the parcel numbers (identifiers),

- the borders of utilization, and
- the fixed border points with their numbers.

The second pillar is the land register. It publishes the data on civil rights on parcels including property rights, easements, and encumbrances. The land register is public. Both, the public register and the cadastre are combined in a parcel database (GDB). Everybody can access the parcel information via local surveying departments, local courts or the Internet. The parcel database is managed by the federal office of metrology and surveying (BEV) in cooperation with the Ministry of Justice. The data in both systems – digital cadastral map (DKM) and in the parcel database (GDB) – are constantly kept consistent. [BEV10]

The model of the digital cadastral map was extended with attributes from the parcel database to fulfil the requirements of the INSPIRE specified data model. The local source model used for the transformation consists of 3 classes (compare Figure 4):

- Grundstueck = parcel
- Grundstueck_EZ = parcel_basic property unit
- KG = cadastral zoning

Parcel (Grundstueck)

The table “Parcel” includes information/data of parcels from the digital cadastral map of Austria. The geometry is defined as polygon. In the data model the identification number of the cadastral zoning (*KG_NUMMER*) is important. This number is a unique identifier in Austria. The parcel can be uniquely identified by the number of the cadastral zoning and the number of the parcel (*KG_GNR*). The coordinates of a reference point in the parcel is defined in *RECHTS-* und *HOCHWERT*. The administrative unit number is also a unique identifier consisting of country number, zoning number, and unique local zoning number.

parcel_basic property unit (Grundstueck_EZ)

This table was generated based on the GDB. It includes all non-spatial attributive data of a parcel. Every parcel can be identified by a number which is presented in the attribute *Nummer*. The area is computed from the results of the survey. This legally valid area is stored in *Flaeche_Gst*. The land register is organized such that each local court manages the part dealing with the parcels in the corresponding court district. The identification of the responsible court is stored in the attribute *Grundbuch*. The basic property unit is defined in *Einlage* and is the smallest unit of property in Austria. There are also more information about the parcel like unique identifier mentioned above e.g. cadastral zoning number and parcel number *KG_NR*, *KG_GNR*. The attribute *GB_EZ* is a unique identifier, which is a combination of land register number and basic property unit.

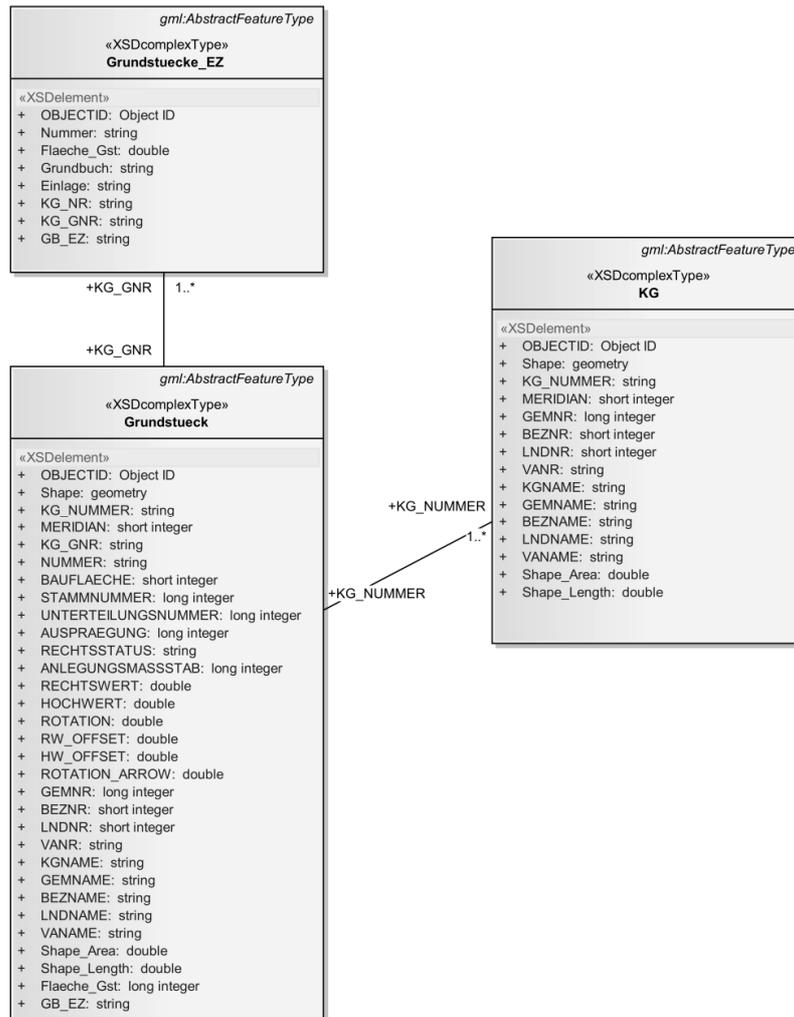


Figure 4: Source model

cadastral zoning (KG_14168)

This table includes all information about the cadastral zoning. The cadastral zoning is the smallest administrative unit in the hierarchical subdivision of Austria.

5. INSPIRE specified target model for Cadastral Parcels

The target model is the Cadastral Parcel model which is defined in Annex I of the INSPIRE directive. Cadastral Parcel is defined in Annex I because many data themes of Annex II and III are based on this data.

The term Cadastral Parcels is defined within INSPIRE as:

"Areas defined by cadastral registers or equivalent" [INS11]

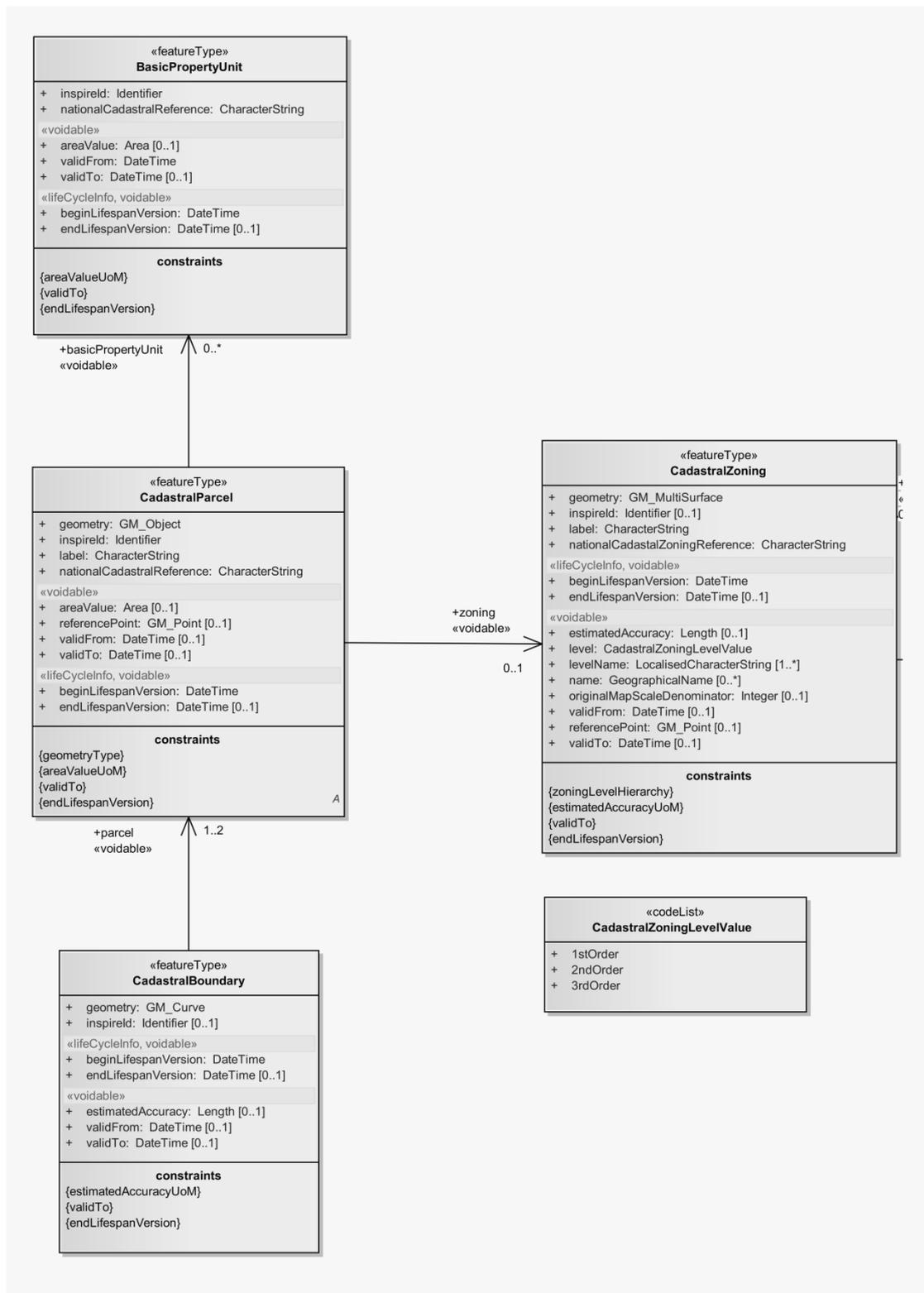


Figure 5: Target model

The target model consists of 4 classes (compare Figure 5):

- BasicPropertyUnit
- CadastralBoundary
- CadastralParcel
- CadastralZoning

6. Implementation of the prototype semantic model transformation

After research about possible tools the prototype transformation has been done with the Humboldt Alignment Editor (HALE). It is an open source tool and was implemented within an EU Project dealing with geodata harmonisation. The project has been coordinated under the guidance of "Fraunhofer Institut für Geografische Datenverarbeitung" (IGD). HALE can import GML Files using versions 2.1, 3.1, and 3.2. It is also possible to import shape files and to upload schemas via Web Feature Services. The imported schemas are visualized in tree structures. The transformation itself is described with OML (Ontology Mapping Language) [Rei10]. The project file is automatically generated as XML file. The result of the transformation can then be exported in form of GML files [Hum11].

In the Technical Guidance for the INSPIRE Transformation Network Service the different transformation tools has been investigated by Matthew Beare and his team. Within this investigation certain requirements have been defined, which have to be fulfilled by the tools for using them [Bea10b].

All recommended tools in this study have been fulfilling the requirements. The decision for using HALE was first of all because it got really good results in the prototype paper of INSPIRE and secondly because it is additional an open source tool [Bea10a].

As source data a cadastral zoning (the municipality Umbach in Lower Austria) has been chosen. Because there was no information available about the performance of HALE this small cadastral zoning had been selected.



Figure 6: Cadastral Zoning Umbach

The source data that has been provided by the Federal Office of Surveying and Metrology of Austria contained shape files of the specified area of interest within the digital cadastral map. This shape files has been imported into ArcGIS and saved into a Personal Geodatabase (.mdb). Because there was still some information missing, additional information (in form of an ASCII file) of the land register had been added to the Geodatabase. However, INSPIRE does not allow the use of slashes. However, they are an integral part of the parcel identification system in Austria. So the slashes used in the cadastral parcel number had to be replaced by hyphen. Afterwards the source data model has been transformed in the Coordinate reference system ETRS89 required by INSPIRE. The coordinate transformation has been done with ESRI Software ArcGIS. The coordinate system of the source model (EPSG-Code 31255) has been transformed into the coordinate system of the target model (EPSG-Code 25833) by using the transformation parameters defined in EPSG-Code 1619. The database has been translated into a GML and a XSD File by the Software FME Translator [FME10]. Finally this XSD File has been imported into HALE as the source schema and the GML file as the source data. The target model has been imported as XSD File.

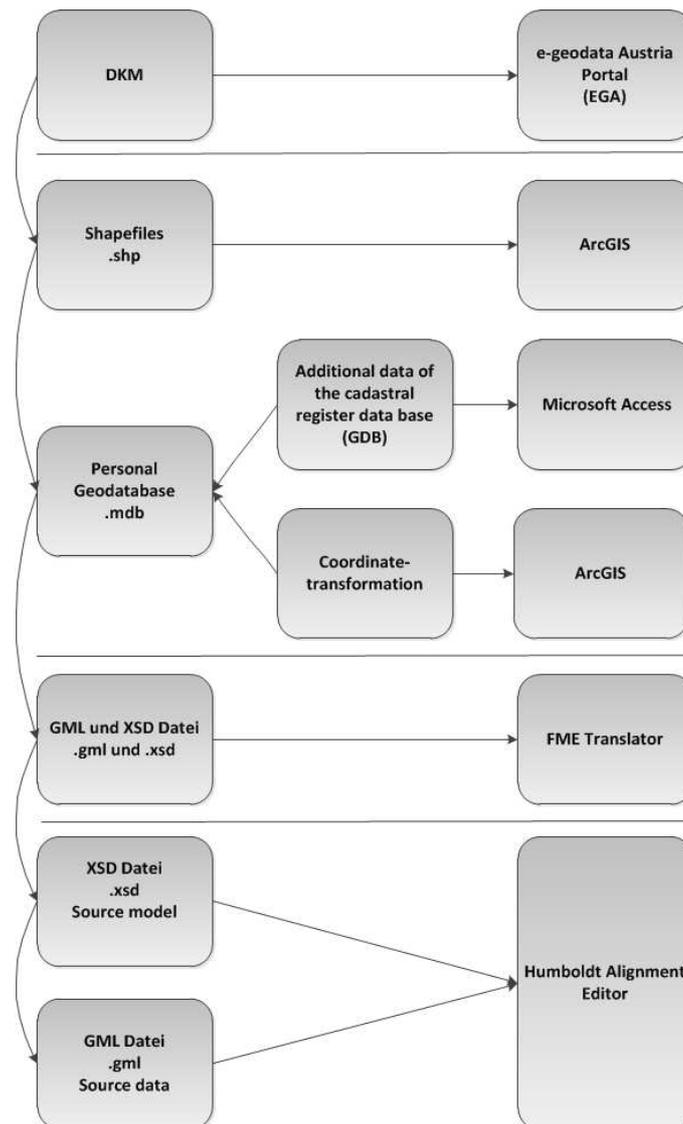


Figure 7: Prototype Transformation Process

Because the recommended transformation tool HALE is still under development, not all necessary functions have been implemented yet. It was not possible to connect a class in the source model to another class such that the information can be used in both classes. Thus the already existing connections between classes in the source model (GB_EZ and KG_GNR) couldn't be used. Therefore, the source model had to be modified. This should typically be avoided but in this case it was indispensable.

The semantic model transformation was done using an overview table. In this table the mapping between both models was defined. It has been visualized, which attribute in the source model is mapped to which attribute in the target model. The table also shows, if the attribute is mandatory or optional. An extra column shows the transformation rule used. This detailed preparatory work allowed a fast and problem-free implementation of the actual mapping rules.

7. Limits of the Implementation

The attributes *beginLifespanVersion*, *endLifespanVersion*, *validFrom*, and *validTo* could not be filled with information because there was no information available in the source model. Therefore they have been declared as INSPIRE Nil Reason.

The attribute *inspire_id* starts with the country code, defined by ISO 3166. For Austria this code is "AT". As second part of the id the name of the data provider (BEV) was chosen because there was no unique number defined for the data providers in Austria at this time of the prototype transformation. The third part of the id consists of the INSPIRE date theme (CadastralParcels) and the object type (CadastralZoning). The object type is recognized and added automatically by the transformation tool. But for the real implementation also this product name will be replaced with a unique number.

The attribute *estimatedAccuracy* has been defined with the default value of 20m. This default value is giving the maximum positional error of the cadastre in Austria within the Alps. When using default values in this prototype the values has been assumed by the values published in the INSPIRE Metadata.

The attribute *originalMapScaleDenominator*, which is describing the scale of the cadastral mapping, was set to the default value of 1:2000 which is also published in the Metadata.

The attribute reference Point, which is the reference point within the polygon of the cadastral parcel, was generated in the class CadastralParcels with the attributes *Hoch- und Rechtswert*. Unfortunately this was not possible in the class CadastralZoning because there the information is missing in the source model. Thus the reference point was generated by the centroid function. This function is calculating the geometrical midpoint based on the geometry. But the problem of this function is that it cannot be used with every shape of geometry because of the different shapes of cadastral zoning the generated reference point can also be generated outside of the cadastral zoning. Figure 8 shows the principal problem based on a simplified visualization. On the left side there is shown a positive example and on the right side there is shown a negative example of the centroid function.



Figure 8: Positive result (left) and negative result (right) of the centroid Function

Because the reference point is not given in the source model it should be added so that this "mistake" can be avoided.

8. Results, Prospect, and Conclusion

Within the study many aspects of an INSPIRE relevant model transformation has been studied. The most relevant results have been:

- » Finding and investigating transformation current tools and their transformation languages they are based on
- » Analyzing the source model
- » Analyzing the target model
- » Establishing, executing, and validating a prototype transformation

It can be concluded that the semantic model transformation is principally possible. The recommended tool, Humboldt Alignment Editor, is still under development and there are still essential functions missing. But HALE is not the only tool for defining such model transformations and there is still a lot of research and implementation going on.

In Austria, a project is currently taking place and should be finished within May 2012 to connect the cadastre and the land register closer to each other. Due to this project also the source data model will be adapted. All mandatory attributes of the INSPIRE target model for cadastral parcel can be fulfilled except the persistent inspire_id for cadastral boundaries, which cannot be included in the used management system.

REFERENCES

- [Aba11] Abart, Günther, Ernst, Julius and Twaroch, Christoph. 2011. *Der Grenzkataster*. Neuer wissenschaftlicher Verlag. 2010.
- [Bea10a] Beare, Matthew, Payne, Simon and Sunderland, Richard. 2010. *Prototype Report for the INSPIRE Schema Transformation Network Service*. 2010.
- [Bea10b] Beare, Matthew, et al. 2010. *Development of Technical Guidance for the INSPIRE Transformation Network Service*. 2010.
- [BEV10] BEV. 2010. Bundesamt für Eich- und Vermessungswesen. [Online] 2010. [Zitat vom: 9. 12 2010.] <http://www.bev.gv.at>.
- [Eis10] Eisenhut, Claude and Kutzner, Tatjana. 2010. *Vergleichende Untersuchung zur Modellierung und Modelltransformation in der Region Bodensee im Kontext von INSPIRE*. München : Technische Universität München, 2010.
- [FME10] FME. 2010. FME Server. [Online] 2010. [Zitat vom: 17. 11 2010.] <http://www.safe.com>.
- [Fra06] Frank, Andrew U. 2006. *Practical Geometry - The Mathematics for Geographic Information Systems. Skriptum zur Vorlesung GIS Theory*. Technische Universität Wien: Institut für Geoinformation und Kartographie, 2006.
- [Hum11] Humboldt. 2011. Humboldt Community Project. [Online] 2011. [Zitat vom: 12. 1 2011.] <http://www.community.esdi-humboldt.eu>.
- [INS11] INSPIRE. 2011. INSPIRE. [Online] 2011. [Zitat vom: 11. 1 2011.] <http://inspire.jrc.ec.europa.eu>.
- [INS09] INSPIRE Thematic Working Group Cadastral Parcels. 2009. *D2.8.1.6 INSPIRE Data Specification on Cadastral Parcels - Guidelines*. 2009.
- [ISO101] ISO 19101. 2002. Norm ISO 19101 Geographic Information - Reference model. 2002.
- [Rei10] Reitz, Thorsten and Templer, Simon. 2010. Humboldt Alignment Editor Manual. s.l. : <http://www.community.esdi-humboldt.eu>, 2010.
- [Sta09] Staub, Peter. 2009. *Über das Potenzial und die Grenzen der semantischen Interoperabilität von Geodaten*, Dissertation. s.l. : ETH Zürich, 2009.

BIOGRAPHICAL NOTES

CONTACTS

DI Eva-Maria Unger
Vienna University of Technology
Gusshausstr. 27-29
1040 Vienna
AUSTRIA
Tel. +43(660)4036814
Email: unger_eva@hotmail.com

PD DI Dr. Gerhard Navratil
Vienna University of Technology
Gusshausstr. 27-29/127
1040 Vienna
AUSTRIA
Tel. +43(1)58801/12712
Fax + 43(1)58801/12799
Email: navratil@geoinfo.tuwien.ac.at
Web site: www.geoinfo.tuwien.ac.at

DI Stefan Klotz
Federal Office of Metrology and Surveying
Schiffamtsgasse 1-3
1020 Vienna
AUSTRIA
Tel. +43(1)21110/3604
Email: stefan.klotz@bev.gv.at
Web site: www.bev.gv.at