Modelling Geo Data for Location-based Services

Jörg Roth
Univ. of Applied Sciences Nuremberg, Germany
Joerg.Roth@FH-Nuernberg.de

1 Introduction

Geo data form the natural resource for location-based services. Geo objects describe the world in terms of natural, artificial and virtual entities that cover the earth's surface. Important functions such as displaying maps, maintaining points of interests (POIs) and navigation strongly rely on the quality of geo data. Spatial data can also be used to improve positioning results (using, e.g., *map matching* in car navigation systems).

The traditional field that deals with geo data is land surveying. As nowadays land survey data is stored digitally, it can in principle be used for location-based services. However, the goals of land surveying often do not meet the requirements of location-based services. As a result, the corresponding data often cannot directly be used.

In this paper, we present a geo data model optimized for location-based services. We identified a number of requirements that the geo model should meet, extended the traditional properties of geo data, which are *thematic*, *geometric* and *topological* properties and introduced *structural*, *organizational* and *meta data* properties.

2 Traditional Geo Models – the ATKIS Case

To discuss geo models, we first have to put them into concrete terms:

- We first have to distinguish *geo models* from *location* (*sensor*) *models*. Geo models represent entities in the world, whereas location models represent the position (of, e.g., a mobile user) as a point in space, an area or a probability function. We carefully have to distinguish a static model representation from dynamic position measurement properties, even though mobile objects (such as trains or ships) may be described using both types of attributes. As a third kind, so called *semantic location models* introduce semantics into location information and generate human-readable symbolic locations.
- Second, we carefully have to distinguish geo model *contents* from their *representation*. Often formats such as KML [4], GML [7], G-XML [3] or LandXML [5] are also considered as geo models, but actually they are descriptive languages. Even though the decision for a specific language highly influences the interoperability between communicating components, it does not be a conceptual decision in the first place (as long as the language is expressive enough).

According to Breunig [2], geo data can be divided into a number of geo-objects, each of them representing real geographic entities such as rivers or abstract entities such as city borders. A geo object can be described by an identifier, topological information, its geometry and thematic content. This overall structure is reflected by data from land survey offices as presented in the next section. We analyzed two data sources and exemplarily imported records (some ten thousands) from land survey offices: TIGER/Line files [17] and ATKIS geo data [1]. We describe the AKTIS case in more detail.

ATKIS (*Amtliches Topographisches Kartographisches Informationssystem*, in English: official topographic cartographic information system) provides geospatial data for governmental purposes in Germany. The ATKIS system unites the distributed data of different land survey offices of the German states. Usually, offices using these data have direct access to ATKIS databases, but private persons or companies can purchase parts of the data exported from the database and exported as EDBS data [6].

As a first observation, we could not use the data directly. Even though the data contain expressive geometric information, it has been optimized for printing maps and not for efficient lookup of geo objects as required by location-based services. Objects, e.g., are spread among different records and are cut at borders of the map grid (every km^2 tile). Due to this and more problems presented below, we created an import and postprocessing workflow (fig. 1):

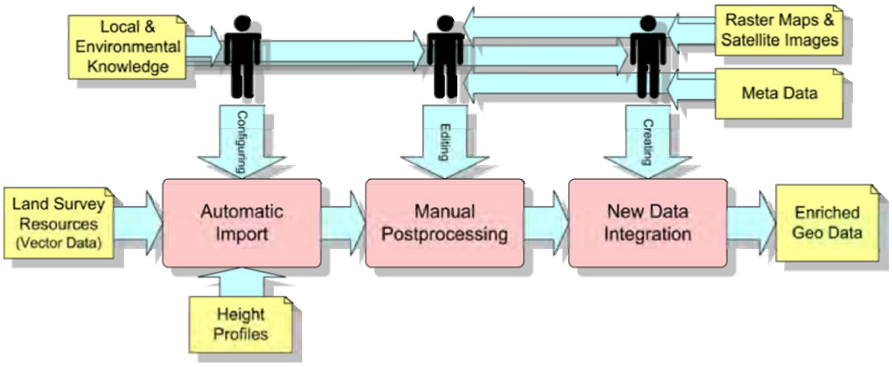

Fig. 1: The import and postprocessing workflow

- An automatic import component parsed and interpreted the original EDBS data format. Related EDBS records have to be merged into single geo objects. Finally, different geometric representations were converted into a single geometric format.
- In a second step (manual postprocessing), we manually merged separated objects that are actually single ones. Typical examples are road objects that are split up at every crossing. This happens because ATKIS prohibits overlapping objects. In addition, we assigned unique names (many objects in ATKIS are unnamed), integrated meta data and classified objects more precisely. The original sources only contain a coarse-grained classification, e.g. universities, schools and kindergartens have the same, meaningless type.
- Third, we integrated new data not contained in the geo data source. E.g., important geo objects such as touristic sights are missing and objects on the same piece of land are often not represented separately. In addition, we integrated *virtual* geo objects such as ZIP zones and GSM cells.

The first step requires manual configuration. Once the configuration is completed, it can be used for similar imports without any new effort. The following steps are more cost intensive as each geo object has to be checked and edited individually. To help an author to perform the second and third step he may use raster data (bitmap maps and satellite images), meta data from different sources (e.g. tourist offices) and should incorporate local and environmental knowledge (which may require visiting geo objects physically).

The experiences of importing geo data from land survey resources can be summarized as follows:

- The import was cost-intensive in three ways. First, geo data is not available for free (in the ATKIS case). Second, development costs arise from the creation of appropriate import filters. Third and most important: cost-intensive manual editing and data integration was required.
- Important thematic information is missing: the classification is not fine-grained enough for locations-based services and some important types of geo-objects are simply missing.

This is because land survey data is mainly collected for property management: In contrast to a location-based service that looks up e.g., a hotel, property management is only interested in any commercial usage of a certain property without further details.

- Geometric information often is inappropriate. Source objects are often artificially cut at the map grid or intersecting objects are divided into different parts. In addition, certain objects are only represented as lines or points without any widths, even though real objects have three dimensions. To perform a test whether a mobile user resides at or in an object, we need at least a two-dimensional ground projection with height levels.
- No or too few meta data is available in the source data. E.g. names are often missing or meaningless and images or links to further resources are missing.
- No or too few topological data is available in the source data. E.g., ATKIS does not express relations between objects and thus topological connections have to be integrated manually.
- Location-based services have additional demands on geo data not fulfilled by the source data. E.g., services may require *distributed* storage and quick retrieval of geo objects. Also improvement of positioning with the help of geo data is currently not supported.

Looking at these drawbacks we introduced a new geo data model that is suitable for innovative services that rely on geo data.

3 Geo Data for Location-based Services – the Hybris Project

Based on the observations above, we want to extend traditional geo data models to be appropriate for location-based services. The *Hybris* framework provides the resulting geo data model with sufficient expressiveness and an authoring environment containing tools for import and editing geo data.

Commercial systems in the area of location-based services often import geo data from hundreds of geo spatial databases [10]. One of or our goals was to represent all required geo information within a single data format. To get the required expressiveness, we first extended the list of traditional properties of geo objects (i.e. geometric, thematic and topological properties). The resulting list of properties (including the traditional ones) is:

- *Thematic properties* mainly provide a classification of geo objects. Useful classifications follow an object-oriented approach that identifies classes and subclasses of objects.
- *Meta data* are names, images, sounds, Web links etc.
- *Geometric properties* describe the geometric shape and are used for rendering maps, distance measurements and testing, whether a given location is at or inside a certain object.
- *Topological properties* are used for street navigation and more complex distance measurements (e.g., distance to get to an object using public transportation).
- *Structural properties* are used to store and access geo data in technical infrastructures. Inside the *Nimbus* project [11-15] we use structural properties to perform a quick geometric lookup of geo objects that are stored in a distributed federation of servers. In contrast to topological properties, structural properties describe relations in terms of storage and access, which are not directly perceived by end-users.
- *Organizational properties* are information about the generation process of geo data. They answer questions such as: What is the data source, who modified an entry and how long is the entry expected to be valid? Such data is important to achieve and preserve the quality of geo data. In contrast to structural properties, organizational properties describe non-technical characteristics.

Even though structural and organizational properties are also meta data in a general meaning, we summarize only those properties that describe content perceivable by an end-user as meta data. Inside the Hybris project, we use an XML language to represent geo objects, whereas each geo object is stored in an individual XML file. That simplifies the distribution of geo objects to different servers. Some details:

Thematic properties: Two fields inside a record provide the object classification: First the ATKIS class imported from the original source provides a rough classification. Second, we introduced a more fine-grained and object-oriented classification with currently 305 classes, that can easily be extended.

Geometric properties: We use a 2.5D representation, i.e. objects are projected on the earth's surface and projected points have an altitude property. We distinguish arbitrary areas presented as *multi-polygons with holes* [16] (fig. 2a), *multi-lines* (fig. 2b) *and point-like areas* (fig. 2c). We always store a multi-polygon representation, also for lines and points, using the so-called *buffer* operation, which computes all points within a certain distance to the geometric object. This simplifies the *location in area test* but still preserves point or line information. We, e.g., can simply check if a car is on a specific road (using the multi-polygon), but we also know the course of the road (using the line information).

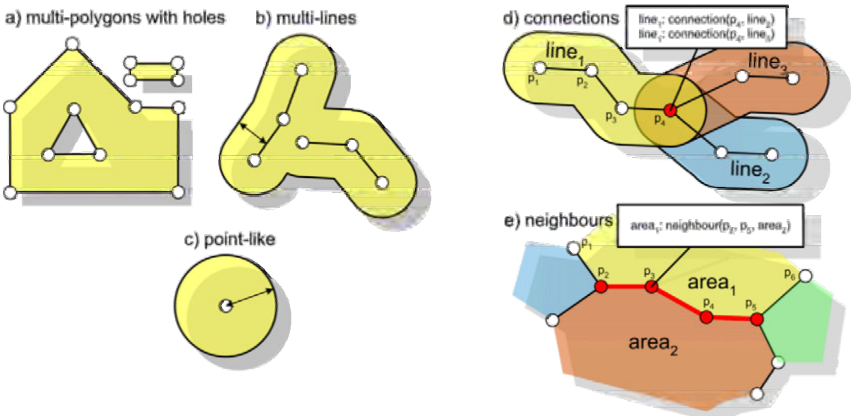

Fig. 2: Geometric and topological properties

Topological properties: Two objects can be topologically connected in one or two dimensions. One-dimensional connections link two objects with the help of a single shared point (fig. 2d). Examples are roads that join in a crossing. Such connections may be unidirectional, to, e.g., express that a road can only be entered from a specific direction. Two-dimensional connections represent common borders, e.g. between neighbouring pieces of land (fig. 2e).

Structural properties: The most important structural property is the *domain name* – a worldwide unique identifier. In contrast to a human readable name as stored inside the meta data, the domain name can be used as a pointer to reference other geo objects (as required to build topological connections). The name in the original sources are often ambiguous, thus we use a hierarchical naming scheme described in [16]. For a fine-grained tuning of the storage mechanism, geo objects additionally have an *area type* (currently one of *stationary*, *mobile*, *temporary* and *abstract*). Mobile areas, e.g., have certain demands on the storage mechanism (e.g. quick update times), thus a server has to know the corresponding characteristics.

Table 1 summarizes the properties and fields in the Hybris geo data model.

Fig. 3 shows the main authoring tool to edit geo objects. The different properties are reflected by different views on the data. The tool allows searching and editing entire geo object trees. It is planned to implement a plug-in interface to build future extensions (e.g. refactoring and rule-checking functions) without affecting the tool core.

Table 1: Hybris geo data properties

Attribute	Description	Values/Example
Structural:		
domain	worldwide unique ID	hierarchical name string / FH_Nuernberg.Woehrd.nuernberg.de
areatype	structural area type	{stationary, mobile, temporary, abstract}
association/ relation	Nimbus association and relation [16]	other domain name / Pegnitz.rivers.geo
Geometric:		
geometry	geometry type	{pointarea, linearea, polyarea}
pointarea	point-like areas	single WGS84 point / C(4927.1782,N,01105.5404,E,300.0m/3)
linearea	lines	multiple line strings / L(4927.1782,N,01105.5404,E,300.0m…)
polyarea	arbitrary 2D areas	multiple polygons with holes / P(4927.1782,N,01105.5404,E…)
lowerbound/ upperbound	2.5D layer	altitude value[m] / 300
captionpoint/ captionangle	point/angle to display the object's name	single WGS84 point 4927.1782,N,01105.5404,E,300.0m angle[°] / 45
Topological:		
connection	lines are connected	tuple (point, other line ID)
neighbour	define a shared border	tuple (startpoint, endpoint, other area ID)
Thematic:		
lsiclass/ atkisclass	object classification	LSI/ATKIS class ID / 22120, 2114
Meta Data:		
name	real object name	UTF-8 string / FH Nürnberg
image	Image file	URL / http://www.fh-nuernberg.de/image.jpg
link	Web link	URL / http://www.fh-nuernberg.de
Organizational:		
expiration	expiration time	duration[min] / 1440
author/modified/ approved	who created/modified /approved the entry	URL or email address / Joerg.Roth@wireless-earth.de

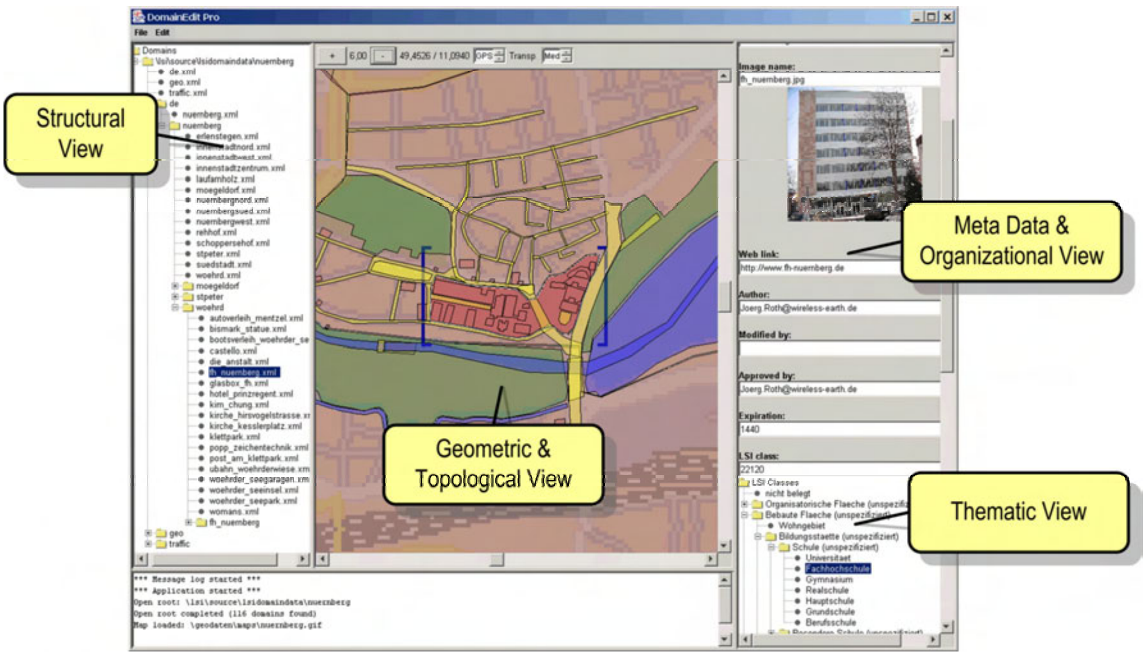

Fig. 3: The Hybris authoring tool

4 Conclusion and Future Work

This paper presents an optimized geo data model for location-based services. Experiences of the import of land survey data have been considered. As an imported observation, geo data of traditional sources have certain drawbacks and cannot directly be used in location-based service scenarios. Import and adaptation is cost-intensive. In particular, structural, organizational and meta data properties have to be added manually.

In the future we want to introduce certain extensions. First, we want to model unsharp borders as necessary to, e.g., model WLAN cells or certain parts of a city (e.g. its centre) which have a fuzzy border rather than a certain geometric limit. The corresponding geometric property has to have a compact representation, but at the same time should allow the execution of efficient geometric lookup operations.

As a second direction we want to generate huge sets of geo data without using land survey sources. In environments such as Wikipedia or Google Earth users make contributions without a forceful central editor. To follow this idea, we want to establish a Web-based workflow that allows arbitrary users to contribute geo data knowledge.

References

1. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV): ATKIS Homepage, http://www.atkis.de/
2. Breunig, M.: *Integration of Spatial Information for Geo-Information Systems,* Lecture Notes in Earth Sciences 61, Springer, 1996
3. Database Promotion Center: *G-XML Project Homepage,* http://www.dpc.jipdec.jp/gxml/contents-e/
4. Google Earth, *KML Specification,* http://earth.google.com
5. Land Development and Transportation Professionals: *LandXML Project Homepage,* http://landxml.org/spec.htm
6. NÖV – Nachrichten aus dem öffentlichen Vermessungsdienst Nordrhein-Westfalen, Vol. 2-3, 1987
7. Open GIS Consortium: *OpenGIS Geography Markup Language (GML) Implementation Specification*, www.opengis.org, 2004
8. Roth, J.: *Mobile Computing*, dpunkt-Verlag, Sept. 2005
9. Roth, J.: *Data Collection*, in Jochen Schiller, Agnès Voisard (eds), *Location-Based Services*, Morgan Kaufmann Publishers, May 2004
10. Roth, J. (ed): 1. GI/ITG KuVS Fachgespräch "Ortsbezogene Anwendungen und Dienste", June 24-25, 2004, Hagen, Informatik Bericht 317, University of Hagen, June 2004
11. Roth, J.: *Semantic Geocast Using a Self-organizing Infrastructure*, Innovative Internet Community Systems (I2CS), Leipzig (Germany), June 19-21, 2003, LNCS 2877, Springer-Verlag, 216-228
12. Roth, J.: *Accessing Location Data in Mobile Environments – the Nimbus Location Model*, Mobile HCI 03 Workshop on Mobile and Ubiquitous Information Access, Udine (Italy), Sept. 8, 2003, LNCS 2954, Springer-Verlag, 256-270
13. Roth, J.: *The Role of Semantic Locations for Mobile Information Access,* Sept. 22, 2005, Bonn (Germany), Proceedings of the 35th Annual GI Conference, Vol. 2, 538-542
14. Roth, J.: *The Distributed Location Resolution Problem and its Efficient Solution*, IADIS International Conference Applied Computing 2006, San Sebastian (Spain), Feb. 25-28, 2006, 3-10
15. Roth, J.: *Detecting Identifiable Areas in Mobile Environments*, The 21st Annual ACM Symposium on Applied Computing, Apr. 23-27, 2006, Dijon, (France), ACM Press, 986-991
16. Roth, J.: *A Decentralized Location Service Providing Semantic Locations*, Habilitation Thesis, University of Hagen, Computer Science Report 323, Jan. 2005
17. U.S. Census Bureau: 108th Congressional Districts Census, TIGER/Line Files, http://www.census.gov /geo/www/tiger/index.html, 2006