INTEGRATION OF SPATIAL DATA WITHIN A GENERIC PLATFORM FOR LOCATION-BASED APPLICATIONS

Steffen Volz, Jan-Martin Bofinger

University of Stuttgart, Institute for Photogrammetry, Geschwister-Scholl-Str. 24D, 70174 Stuttgart - (steffen.volz, jan-martin.bofinger)@ifp.uni-stuttgart.de

Commission IV, WG IV/2

KEY WORDS: Federated Infrastructure, Location-Based Services, Multiple Representations, Spatial Data Integration, XML/GML, Conflation, Matching Conflicts

ABSTRACT:

Within the last few years, a new kind of applications has evolved: location-based services are on their way to become one of the leading powers within the field of information technology. At the University of Stuttgart, a research project called NEXUS has been initiated to develop an open and global platform supporting all possible types of mobile, location-based information systems. A federated infrastructure will be built up that manages a detailed model of the real world, which is called the Augmented World Model. Thus, an integrated view on the distributed data sources and services that have to cooperate within NEXUS as well as a standardized interface can be offered to the applications. In order to realize the generic approach of NEXUS, two crucial prerequisites have to be fulfilled: First of all, different data providers have to be able to integrate their data into the Augmented World Model. For this reason, a schema integration takes place that maps the object classes of existing data models onto the classes of the Augmented World Schema. Second, corresponding object instances within the data of the Augmented World Model must be matched. Thus, adjoining or overlapping data sets can be merged in order to derive one complete and semantically enriched spatial representation meeting the needs of different location-based applications. Therefore, appropriate conflation algorithms are being developed. The paper briefly introduces the idea of NEXUS and describes the architecture and the data model of the platform. Then, the approaches for spatial data integration are explained in detail.

1. INTRODUCTION

Location-based services can be defined as applications which know the positions of their users and provide information and services accordingly on a mobile computer device. Thus, the main criterion for selecting information is the spatial context a user is in. Today, only isolated applications like car navigation systems or mobile tourist guides are offered by different providers. They are restricted to certain data formats, certain mobile communication techniques and certain sensor technologies. In order to provide an open and global platform on top of which various types of location-based applications can be developed, NEXUS aims at integrating all the different components.

The quality of a global location-based information system like NEXUS especially depends on the availability of detailed and manifold spatial data. For this reason, heterogeneous data sources that have been acquired by different providers in different underlying data models and scales must be integrated into the NEXUS platform. In order to achieve this goal, a global data model called the Augmented World Model (AWM) has been developed that contains all objects which are relevant for location-based applications. For the integration of heterogeneous spatial data within the NEXUS system, object classes of existing data models have to be mapped onto the classes of the AWM. Thus, a uniform view on the data as well as a standardized interface can be offered to the applications. As a prerequisite for a further usage of heterogeneous data sources within NEXUS, multiple digital instances of one real world object must be identified. Therefore, matching algorithms are being applied in order to merge overlapping or combine adjoining data sets.

This paper comprises an overview of the NEXUS platform in section 2. Section 3 describes how heterogeneous road data sources can be integrated into the Augmented World Model. In section 4 the conflation approaches within NEXUS are presented.

2. THE NEXUS PLATFORM

The NEXUS platform can be described as an open and distributed environment for location based applications. It consists of a federated architecture which manages services and information resources and provides a consistent view for the applications: the Augmented World Model.

2.1 Architecture of the Infrastructure

The components of the NEXUS infrastructure are organized in three tiers (see Figure 1): applications, federation and services. Basically, the architecture of NEXUS can be compared to the one of the World Wide Web. Just like Web servers, NEXUS data servers can be provided by anyone and only have to be registered in a certain directory that stores the address and the properties of an information resource. NEXUS applications are similar to Web applications like browsers or applets. However, the intermediate tier of the NEXUS architecture, the federation, has no counterpart in the WWW. The federation organizes the response to client queries within the distributed NEXUS architecture.
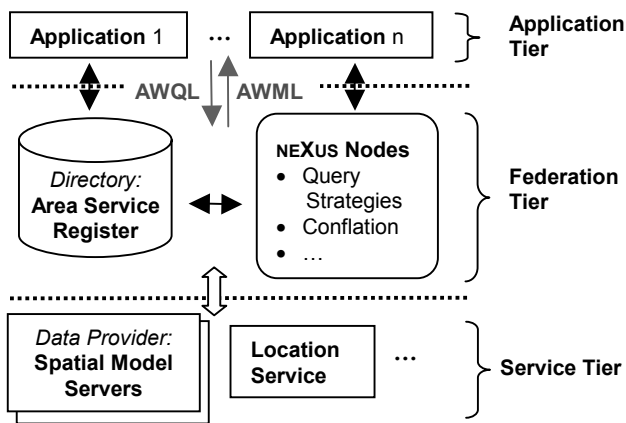
Figure 1: Architecture of the NEXUS platform

2.1.1 Application Tier: NEXUS applications are typically running on small mobile devices which are able to determine their position via appropriate sensors. When they are performing queries, they contact the NEXUS nodes of the federation layer using wireless communication technologies like WLAN or UMTS. Queries are formulated in the XML-based Augmented World Query Language (AWQL). AWQL contains simple spatial and attributive predicates. Usually, the clients receive the results of their queries in the Augmented World Modeling Language (AWML). AWML is also defined in XML and allows to serialize the objects of the AWM.

2.1.2 Federation and Service Tiers: Any application request is directed to a *NEXUS node*. On each node, a federation component is running. The NEXUS nodes are responsible for developing a strategy to respond to application requests. They distribute the queries on the appropriate components of the Service Tier, namely the Spatial Model Servers and the Location Service. A *Spatial Model Server* stores and manages the static spatial data (like buildings, streets, etc.) of one or more geographical areas. It also provides geoprocessing functionalities. The *Location Service* deals with the management of the user positions. Since huge amounts of location data is updated frequently, the properties of standard databases are inadequate. Thus, we have developed a main memory based solution for efficiently storing and retrieving the position data of mobile users (A. Leonhardi & K. Rothermel, 2001).

In order to create a strategy to answer application requests, a NEXUS node must first identify, which spatial data have to be used for query processing. Therefore, it addresses the *Area Service Register* which can be understood as a metadata repository storing general information about spatial data sets (like spatial extent, level of detail, stored object classes etc.) and the addresses of the Spatial Model Servers where they are located. After having queried the appropriate Spatial Model Servers and the Location Service, the NEXUS nodes conflate the data that is returned from the Service Layer and provide additional services like navigation or map production on their own. Eventually, the results are propagated to the application.

The NEXUS infrastructure contains further services which are for example described in (Volz, 2001).

2.2 The Augmented World Model

The Augmented World Model (AWM) is an object-oriented information model for location-based applications (Nicklas & Mitschang, 2001). It allows a unified view for the applications and enables the federation of spatial data from different providers. The object classes of the AWM are defined in the Augmented World Schema (AWS). The AWS comprises the standard class schema, a generic set of all object classes that might be of relevance for location-based applications. To achieve a reasonable semantic, each standard class contains an extensive set of attributes. However, most of the attributes are declared *optional*, i.e. a NEXUS service does not have to make the effort to provide all of the data, but if it wants to, the name and type of the attributes are already defined. In case that an application still needs additional object classes that further specify the existing ones, they can be inherited from the classes of the standard schema, forming a so-called extended class schema (see Figure 2).
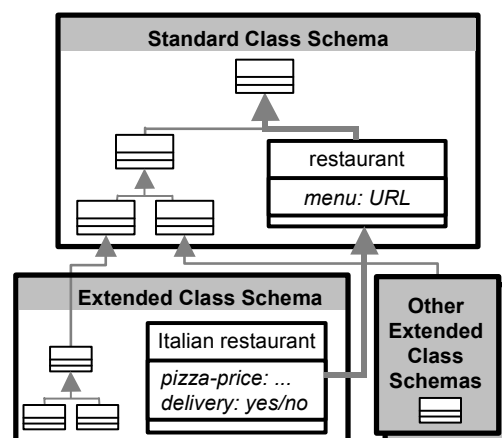

Figure 2: The AWS standard class schema can be extended according to the needs of applications

3. INTEGRATION OF HETEROGENEOUS DATA SOURCES INTO THE AWM

Since the NEXUS platform is aimed at providing an extensive and generic data basis for different kinds of location-based applications, data sources from all possible application fields should be integrated in the AWM. Therefore, existing data models have to be mapped onto the Augmented World Schema. This allows on the one hand a further use of already available data. On the other hand, data from multiple sources can be federated since they are available in a common data format. In the following sections, the process of mapping road data from different sources onto the AWS will be described.

3.1 Existing Data Models for Road Data

The growing popularity of car navigation applications in recent years has lead to a multiple acquisition of road data in different formats. In the following sections the two most common data models for road data in Germany will be introduced. The Geographic Data File (GDF) is used by the major digital road data suppliers, while ATKIS is a topographic and cartographic database. Both formats are acquired in an approximate scale of 1:25 000.

3.1.1   The GDF Data Model: GDF 3.0 (CEN TC 278, 1995) is a European standard that is used to describe and transfer road networks and road related data. It is much more than a GIS data format, because GDF regulates how data should be captured and how features and their relationships have to be defined. Features correspond to "real world objects" like streets and railways or administrative areas. The feature category defines the type of representation of a feature. Objects represented by a point, line or area are called *simple features*, while *complex features* are composed of a group of simple features. Features contain geometrical as well as topological information. Every feature belongs to a specific feature class. The following feature classes are important for road traffic:

- *Junction* (point): feature bounding a Road Element
- *Road Element* (line): smallest independent unit of the road network having a Junction at each end
- *Road* (complex): contains different Road Elements
- *Intersection* (complex): composed by Junctions and Road Elements

Figure 3 illustrates the relations between simple and complex road traffic features.
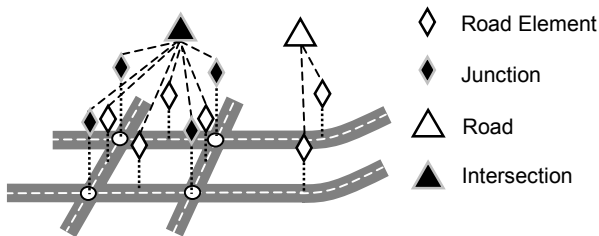

Figure 3: Simple GDF features can be aggregated to complex ones (after Walter, 1997)

Features can also share relationships. Relationships are used to express associations between two or more features like, for example, a prohibited manoeuvre from one road to another.

3.1.2   The ATKIS Data Model: The data model of ATKIS (Authoritative Topographic Cartographic Information System) has been developed by survey offices of the federal states of Germany (AdV, 1998). It is structured into seven functional classes like e.g. settlement, vegetation and traffic. Each functional class is again subdivided into object groups. Concerning traffic, its object groups are road traffic, railway traffic, air traffic, etc. An object group includes several object types. Object types specify the geometrical representation and the attributes of real world objects. In case of road traffic, defined object types are for example roads, squares or ways. Instances of the object types are the ATKIS objects which are built up by object parts. Object parts contain the geometry of ATKIS objects as vector elements. Therefore, each ATKIS object comprises at least one object part. In case of roads, further object parts are established, if the topology (in case of crossings) or the attributes (e.g. the road width) of an object change. ATKIS demands to build complex objects in case that a road has two or more physically seperated lanes.

3.2 Schema Integration: Mapping existing Road Data Models onto the AWS

Concerning the mapping of existing road data models onto the object classes of the Augmented World Schema, it is important to preserve the original properties of features as far as possible.

First of all, a suitable amount of object classes has to be defined in the AWS and thereupon the relations between them and the corresponding object classes of GDF and ATKIS must be determined. Afterwards, the XML schema definition of the Augmented World Modeling Language (AWML) can be extended to road traffic objects and thus the original GDF and ATKIS formats can be converted into AWML.

3.2.1   Defining the object classes: On the basis of the two underlying data models GDF and ATKIS, a generic set of object classes has been established in the AWS that allows to represent all object classes of the existing data models (see Figure 4). They are all derived from the abstract class *StaticObject* which is the superclass of all static spatial objects (like buildings, streets, etc.) within the AWS. To each object class of the AWS, necessary attributes have been assigned. Only some most basic of them were declared mandatory.
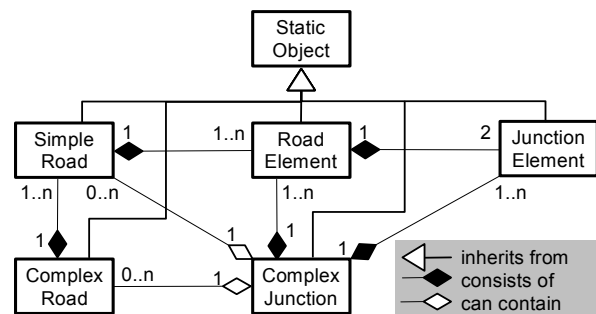

Figure 4: UML representation of road objects in the AWS including cardinalities

Basically, there are 1:1 relations between the object classes of GDF and AWS. However in GDF, Road Elements which are sharing the same street name do not need to be combined to form a road feature. But within the AWS, *RoadElements* carrying the same name have to be aggregated to a *SimpleRoad* object. Additionally - in case of a multi-lane road - GDF road features that have the same name are merged into an AWS *ComplexRoad* (see Table 1).

GDF	Relation	AWS
Junction	1:1	*JunctionElement*
Road Element	1:1	*RoadElement*
Road Element	*n*:1	*SimpleRoad*
Road	1:1	*SimpleRoad*
Road	*n*:1	*ComplexRoad*
Intersection	1:1	*ComplexJunction*

Table 1: Relations between object classes of GDF and AWS

ATKIS	Relation	AWS
Object part	1:2	*JunctionElement*
Object part	1:1	*RoadElement*
Object	1:1	*SimpleRoad*
Complex Object	1:1	*ComplexRoad*
Not existing	–	*ComplexJunction*

Table 2: Relations between object classes of ATKIS and AWS

Regarding the relationships between ATKIS and AWS object classes (see Table 2), it can be recognized that each object part corresponds to exactly one *RoadElement* being bounded by two *JunctionElements*. Except the AWS-*ComplexJunction* class

which is not existing in ATKIS, all other object classes share 1:1 relations.

Figure 5 illustrates the mapping of data from GDF and ATKIS onto the Augmented World Schema.
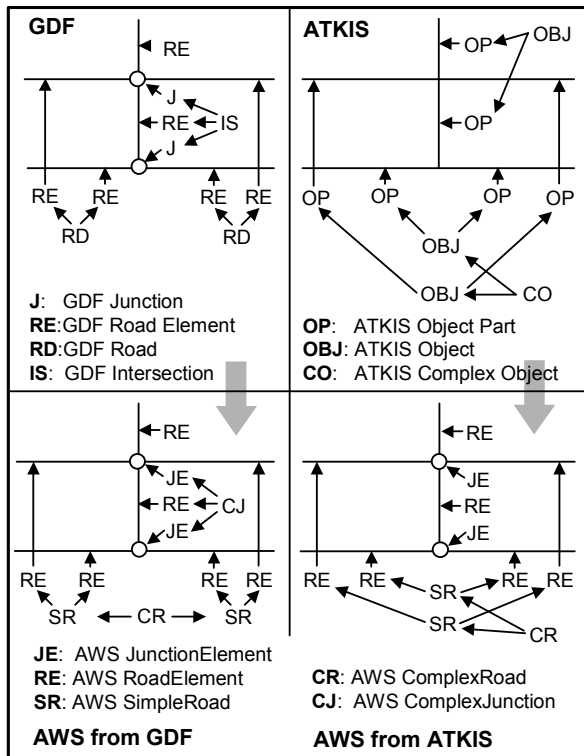

Figure 5: Representation of GDF and ATKIS in the AWS

3.2.2   Developing the AWML Schema: After having built up the object classes for road data within the Augmented World Schema, the definition of the Augmented World Modeling Language (AWML) took place applying the XML schema language (W3C, 2002). Since NEXUS is aimed at providing an open platform, standards are adapted wherever possible. For the geometrical representation of the Augmented World objects, AWML implements the geometry schema definition of the Geography Markup Language developed by the OpenGIS Consortium (OpenGIS, 2001).

Each AWML file consists of a header and a data section. The header carries meta information about the file like its spatial extent, contained object classes, information about the origin of the data, etc. The data section includes the XML description of the AWS objects. Objects of superior classes are referencing subordinate instances by the *href* attribute defined in the XLinks specification (W3C, 2002). For example, a *RoadElement* contains XPointers to the IDs of its From- and To-*JunctionElements* within its *href* attribute. XPointers enable referencing individual elements within an XML document.

3.2.3   Converting GDF and ATKIS into AWML: This section describes the conversion of existing road data into AWML files. Both GDF and ATKIS data sources are stored in a sequential interchange file format requiring special parsers to extract the necessary information. Implementing such parsers can be a tedious task, especially if you are not familiar with the file format. Since AWML is an XML language, the generation

of AWML is relatively easy to conduct, because many software libraries support handling XML data with standard interfaces like DOM (Document Object Model). The following paragraphs will give a brief description of the conversion process in our work.

GDF: GDF data are stored in ASCII-based interchange files. In order to extract data of a specific region out of a GDF file, a preprocessing step is performed. It cuts out any feature contained in or intersecting the minimum bounding rectangle of the area in focus. After this preprocessing the conversion tool starts building road objects by parsing the extracted GDF records and inserts them into the Document Object Model (DOM) of an AWML document. As mentioned in section 3.2.1 the tool has to construct *SimpleRoads* out of GDF Road Elements having the same street name and/or number, because in AWS the official name and number of a road are stored in the *SimpleRoad* and not in the *RoadElement*.

The conversion tool is written in C++ and uses Xerces from the Apache XML Project (Apache, 2002) to generate the DOM of the AWML file. The DOM is automatically validated against the XML schema of AWML.

ATKIS: ATKIS data are typically exchanged via an ASCII-based format called Uniform DataBase Interface (EDBS). This format has been developed almost twenty years ago and focuses on a minimization of the data volume by a line-based ordering. Thus, it does not preserve any object structures, but the information belonging to one object is scattered throughout the file. To allow an easier parsing, EDBS has been converted into an object-structured file representation out of which the AWML file is generated.

The parser itself has been written in Java. It reads the ATKIS objects and stores them in appropriate containers according to the object class they belong to. The containers are then written into an AWML file using the JDOM library (JDOM, 2001). JDOM contains classes for generating and parsing DOM instances.

4.   CONFLATION OF SPATIAL DATA IN NEXUS

In order to realize a co-processing of heterogeneous spatial data from different origin within NEXUS, it is not only necessary to integrate them in the Augmented World Model. Rather, multiple digital representations of one and the same real world object have to be identified to enable the conflation of adjoining and overlapping spatial data sets. Thus, coherent and semantically enriched data can be generated that best meet the needs of different types of location-based applications.

Generally, a conflation process is the more successful, the less discrepancies between two spatial representations in terms of scale and data model exist. In NEXUS, not only data with a high degree of similarity must be matched, but also correspondencies between heavily differing objects must be detected. The following sections describe the approaches for the conflation of spatial data within the NEXUS platform.

4.1   The Conflation Process

Each time a client query that implies a usage of heterogeneous spatial data sources is directed to a NEXUS node, a conflation process has to be performed. Then, further services can operate on the conflated data (see Figure 6).
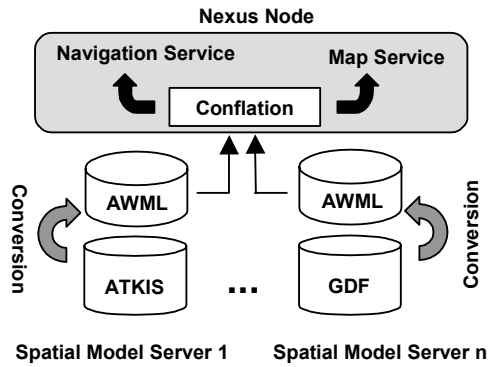
Figure 6: The role of conflation within NEXUS

Since the conflation process must be carried out fully automatically, the algorithms have to be robust and fault-tolerant in order to guarantee an acceptable quality of service for the user. Therefore, errors of the conflation procedure have to be detected by the system itself. This can be realized by introducing quality measures expressing the success of a matching operation between two spatial features. With respect to user acceptance, also the time response of the process has to be considered, which, of course, depends on the amount of data that have to be matched. Altogether, the conflation process within NEXUS consists of four steps:

a. **Schema integration:** Generally, schema integration comprises the development of a database schema that allows to represent objects from different origin (Devogele et al., 1998). In the case of NEXUS, this step conforms to the mapping of heterogeneous data sources within the AWS that was described in section 3.2.

b. **Analysis of meta data:** Each AWML document contains meta information about its data. For example, the coordinate system, the data origin, the spatial extent, etc. are specified in the header section. This information can be exploited to determine, which operations have to be carried out in the conflation process, e.g. if a coordinate transformation has to be done and so on.

c. **Determination of the conflation category:** The conflation category is defined by the type of query the user performs. Either, any data of two AWML files are conflated in overlapping areas to generate one semantically enriched file or two AWML files are merged at their boundaries to receive one coherent data set.

d. **Matching of object instances:** Using semantical, geometrical and topological indicators, object instances are compared in order to determine their degree of similarity. The matching process generally involves many conflicts (see section 4.2) which not only come up due to different data models or scales, but also because of different times of acquisition or erroneous data capturing. Therefore, in some cases a matching process is ambiguous or doesn't yield any result at all.

4.2 Potential Matching Conflicts

In the following sections, some reasons for matching conflicts are outlined, especially considering road data.

4.2.1 General conflict situations: Depending on the diversity of spatial data sets, different conflicts in the matching process arise. In particular, problems occur when data sets which are represented in different data models have to be merged. For example, if corresponding objects are described by lines in one data set and polygons in another data set, it is not possible to conduct the matching without preprocessing. Therefore, in order to work on common geometrical elements, polygons would have to be reduced to lines by a skeleton algorithm accepting some loss of information.

Another problem is the different levels of detail in which real world objects are stored in data sets. The generalization process takes place during the acquisition of data by an operator and can also be done in order to adapt objects to according scales. Considering this fact, the cardinality of matching objects is not limited to one to one, even many to many relations are possible, considerably increasing the complexity of the conflation process.

As mentioned before, non-geometrical properties of object instances also contribute to the matching process. Therefore it is necessary to identify corresponding attributes in the schema integration step. Even if the identification is successful, there can be discrepancies between common attributes due to the fact that data are collected on the basis of different data sources.

4.2.2 Conflicts with road data: In order to become aware of possible conflict situations in the matching process, both the AWML representations of ATKIS and GDF road networks have to be superimposed and visualized. Since AWML is an XML-based language, it is straightforward to convert AWML files into SVG files (Scalable Vector Graphics) using XSL Transformations (W3C, 2002). In addition, interactive features for further examinations like the selection of objects have been implemented within a GUI using JavaScript.
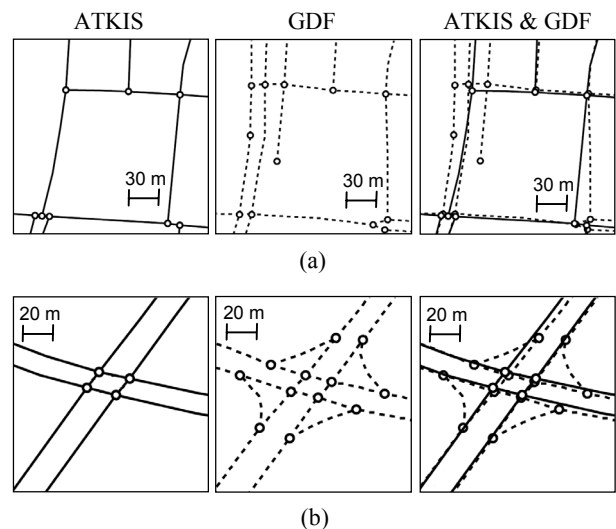

Figure 7: Conflict situations while matching road data of GDF and ATKIS (after Walter, 1997), visualized in SVG

Figure 7 shows typical conflict situations in the matching of road network elements. Because GDF is explicitly meant for car navigation, it is sometimes more detailed than ATKIS. For example, a road represented by one lane in ATKIS data can be represented by a multiply digitised road in GDF (see (a)). Also,

complex intersections are captured with a higher accuracy in GDF than in ATKIS (see (b)).

4.3 Matching Approach for Road Data

Objects within geographical information systems have a semantical, a geometrical and a topological component. The semantical and geometrical properties are used as indicators when comparing two digital representations of the same real world object during a feature-based matching process. In addition to the feature-based approach, the relational matching also takes into account the relations of an object to other objects within the same data set. For example, topological relations like "is contained in" or "has *n* neighbours" are considered.

In a matching process, a set of objects A is mapped onto a set of objects B by an appropriate function. Walter (1997) has presented a relational matching solution for ATKIS and GDF road data. First of all, a list of potential matching pairs is generated that can have cardinalities from 1:0 up to n:m. Then, the pairs are rated by feature-based and relational indicators. Thus, the list can be continuously refined by eliminating improbable assignments. In order to achieve a quality measure of the matching process, the result of the automatic procedure is compared to a manually derived reference list. In principle, the same approach can be adopted for the matching of road data within NEXUS.

Since GDF and ATKIS can be represented in AWML, XSL Transformation (XSLT) (W3C, 2002) will be applied for the implementation of basic matching algorithms like the comparison of line lengths or object angles. Still, the capabilities of XSLT concerning more sophisticated conflation operations have to be analyzed. Then, it can be decided which functionalities need to be implemented using external programming languages like Java.

5. CONCLUSION AND OUTLOOK

Within this paper, the NEXUS platform - an open, global infrastructure for location-based applications - and its generic data model - the Augmented World Model - have been presented. It has been shown, how heterogeneous data sources can be integrated into the Augmented World Model. Furthermore, the concept for conflating spatial data within NEXUS was described.

Future research will have to focus on an extension of the conflation approach, so that not only data of similar scale and data model but also strongly differing resources can be further processed within NEXUS. Therefore, a generic conflation model is going to be developed that aims at reflecting a maximum of possible conflict situations between heterogeneous spatial data. Furthermore, it will be examined how the time response of conflation operations can be optimized by appropriate parallelization methods.

6. ACKNOWLEDGEMENTS

7. REFERENCES

Apache XML Project, 2002. http://xml.apache.org

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 1998. Amtlich Topographisch-Kartographisches Informationssystem (ATKIS). Bonn.

CEN TC 287, 1995. Geographic Data Files Version 3.0, GDF for Road Traffic and Transport Telematics.

Devogele, T., Parent, C. and Spaccapietra, S., 1998. On spatial database integration. *Int. J. Geographical Information Science*, 12 (4), pp. 335-352.

JDOM, 2001. http://www.jdom.org/

Leonhardi, A., Rothermel, K., 2001. Architecture of a Large-scale Location Service. Technical Report 2001/01. Department of Computer Science. University of Stuttgart. 1-17.

Nicklas, D., Mitschang, B., 2001. The NEXUS Augmented World Model: An Extensible Approach for Mobile, Spatially-Aware Applications. *Proc. of the 7th Int. Conf. on Object-Oriented Information Systems*, Calgary.

OpenGIS Consortium (OGC), 2001. Geography Markup Language (GML) 2.0 Implementation Specification. http://www.opengis.net/gml/01-029/GML2.html (accessed 11-17-01)

Volz, S. 2001. Information Management for Location Aware Applications, in: D. Fritsch & R. Spiller, eds, *'Photogrammetric Week '01'*, Heidelberg, pp. 5–12.

Walter, V. and Fritsch, D., 1999. Matching Spatial Data Sets: a Statistical Approach, *International Journal of Geographical Information Science* 13(5), 445–473.

Walter, V., 1997. Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF. Dissertation, Deutsche Geodätische Kommission (DGK), Reihe C, Heft Nr. 480.

World Wide Web Consortium (W3C), 2002. http://www.w3.org