SEMANTIC AND GEOMETRIC INTEGRATION OF GEOSCIENTIFIC DATA SETS WITH ATKIS – APPLIED TO GEO-OBJECTS FROM GEOLOGY AND SOIL SCIENCE

Guido von Goesseln, Monika Sester

Institute of Cartography and Geoinformatics, University of Hanover, Germany,
(guido.vongoesseln, monika.sester)@ikg.uni-hannover.de

Commission IV, WG IV/7

KEY WORDS: **Cartography, Geology, Soil, Integration, Change Detection, Update, Correction**

ABSTRACT:

Solving problems in an environmental or geoscientific domain usually involves various data from different sources: Typically, not all the data is available in one dataset, but is distributed, has different formats, and different thematic focus. Data integration is therefore needed for the combination and analysis of different data sets. The project "Geotechnologien – New methods for semantic and geometric integration of geoscientific data sets with ATKIS – applied to geo-objects from geology and soil science" funded by the German Ministry of Research investigates different aspects of data integration and aims at developing methods for automatic integration from different data-sets and their consistent management in a federated data base. Three different vector data sets are used in the project: the topographic data set ATKIS, the geological map and the soil science map. The topographic component of the geological and the soil science map is based on topographic maps, but the frequency of update differs between ATKIS and the geoscience maps. Thus, the data sets differ not only in thematic content, but also referring to seemingly identical topographic objects. In the first step of the project the objects and their semantics in the different data sets are investigated in detail, the object water will be the first very promising candidate for identifying the conformity between the different data sets.
In the second step, a geometric reconciliation will be established. To this end, firstly corresponding objects will be identified using matching procedures. The identified objects from the geoscientific maps will be adapted to the reference objects from ATKIS. Following this procedure the automatic change detection shows all the differences between the reference data set and the data sets to be integrated.

1. INTRODUCTION

Geo-information systems (GIS) allow for an integration of all kinds of spatially referenced digital data sets. However, when simply superimposing data sets of different origin and/or thematic focus, discrepancies show up, especially geometric and semantic differences between seemingly similar objects.
Such differences might be surprising when the data sets are originally derived from the same source, namely the earth topography. These differences can be explained by the fact that the data stems from different sensors, is acquired based on different data models, scales or human operators.
The heterogeneity of the data sets makes it complicated and sometimes even impossible to handle these data of unequal representation and quality in a homogenous form. Therefore, harmonisation and data integration is needed in order to facilitate the usage of the data for various tasks, allowing the derivation of new thematic information, the automatic verification regarding their quality, and improving the accuracy.
Additional benefits of the integration is the reduction of financial and human resources. Data which is not available in one source can be derived from another data set. In our example, the aim is to propagate updates from a reference data set (in this case the digital topographic data set ATKIS) to geoscientific data sets, that have originally been acquired based on topographic maps. The "Geotechnologien"-project "New methods for semantic and geometric integration of geoscientific data sets with ATKIS – applied to geo-objects from geology and soil-science", which is funded by the German Ministry of Research and Education, is divided into three sub-projects (Sester et al., 2003). The main task of the institute of Cartography and Geoinformatics is the automatic detection of discrepancies and changes, and the subsequent adaptation of the updated dataset. The two other partners in the project focus on the representation of the different data sets in a federated data base on the one hand, and on the integration of vector and image data on the other hand.
The paper is organized as follows: After a short presentation of related work, the different data sets are described. The focus is set on the main differences between the data sets, e.g. the time of updating, the data-model and the thematic content. This is followed by the description of the semantic relations between the data sets. Then the geometric integration and adaptation is presented with the description of the implemented iterative closest point (ICP) algorithm. After the integration, a change detection has to be performed in terms of an intersection and analysis. The first results of the project achieved so far concerning test objects "water" are described in this paper. The presentation concludes with open questions and an outlook on the tasks which are planned to be done in the near future.

2. RELATED WORK

Data integration is being investigated with different focus: on the one hand, data of different sources is integrated for a common data analysis in order to derive new knowledge. Secondly, data can be integrated and fused for mutual benefit: (Walter, 1997) presents an approach that fuses two different data sets with road information with the aim of mutually exchanging attributes of the two data sets involved. The integration of vector data and raster data is being investigated in the partner project with the aim of enriching a 2D-vector data set with 3D-information (Butenuth & Heipke, 2003). It is also popular in the domain of 3D-city modelling, where 2D-groundplans from buildings and 3D-surface models are fused (e.g. (Brenner, 2000)). Data integration or data matching is also needed for update purposes, e.g. when a data provider has to deliver up-to-date information details to his customers (Badard, 1999). Integration can be used for data registration, when one data set is spatially referenced, and the other has to be aligned to it (Sester et al., 1998). Finally, data integration is needed for the generation of Multiple Resolution Data Bases (MRDB); in this case objects of different geometric and thematic resolution have to be fused (Mantel, 2002).

3. DATA SETS USED

Three different data sets are handled in this project. The german digital topographic map (ATKIS) will be taken as the reference data-set because of its actuality and very accurate geometry. The geological map and the soil-science map are delivered by the external partners of the project. The topography or at least parts of it are represented in all three maps, but so far hardly comparable. The visible discrepancies between the topographic representations can be explained by the way the data sets were created.

3.1 Digital Topographic Data Set – ATKIS

German National Mapping Agencies are responsible for the timely acquisition and homogenous representation of the topography in digital data sets and maps. Objects from the real world are classified using seven object classes (e.g. vegetation, settlement, water). Each of these classes is sub-divided into object groups, containing the different kind of objects (e.g. sea, ocean, river). The update period of the object classes varies from one year up to three months (high-actuality) for streets and important road-connections (LGN 2003).

3.2 Digital Geological Map – GK25

In the state of Lower Saxony the "Niedersächsisches Landesamt für Bodenforschung (NLfB)" (soil science institute) is responsible for the creation of the geological map in scale of 1:25.000 (GK 25). The purpose of the GK is to provide comprehensive and detailed information about the geological structure, representing the different levels from earth surface to a depth of two meters. Representative geological cuts and information about the stratigraphic sequence up to hundreds of meters enrich the thematic content of the geological map. Soil science, hydrogeology, engineer-geology and deposit geology use this map as base material. The creation of the geological map is based on drills down to two meters, the results are integrated to the map. The drills are placed with a point-to-point distance of 100 meters. Drills down to some hundred meters are used for the research of structures in deeper layers. Based on these punctual information models of areal-data are derived (Ad-hoc 2002).

This geological information is related to the topography (i.e. a topographic map) atf the point of time the geological map is created. Later on, only thematic updates of the geological content are integrated without updating topographic objects. Only when new information about the geology in singular areas has been acquired (through drills or other sources) this will be integrated into the geological map and then additionally also the topographic content of that area might be updated.

3.3 Digital Soil-Science Map – BK25

Mapping the soil surface is done at the NLfB, with the soil-science map as product showing the extension of different soil-types classified according to structure and characteristics. The content of the soil-science map is built up by the acquisition and characterisation of soil-areas with drills and soil-explorations. The soil science map shows the structure down to a depth of two meters, covering different types of attributes, e.g. soil systematic classification, soil texture, soil horizons. Appropriately classified areas are gathered to form soil units. The soil science map is used for various applications in different sciences like engineering, hydrogeology, biology, history and archaeology. In order to create a soil-science map existing information is gathered in a concept-soil-map including punctual drill results, different kinds of geoscientific maps, as well as topographic maps and images from airborne photogrammetry and remote sensing.

Instead of using a strict and regular raster, variations of relief, vegetation and soil humidity provide possible locations for drills to be taken. The punctual information is used to extrapolate the areal extension of the soil-textures which are represented as 2D-polygons, e.g. mapping units (Ad-hoc, 1994).

Topographic elements like water-areas are taken from topographic maps and integrated into the concept maps. In most of the cases this is done manually, by copying the elements from on map to the other. The actuality of the topographic elements depends on the creation of the soil-science map. Similar to the geological map, the changing of the thematic content of the soil-science map is not comparable to the changes which happen to the topography. From the moment the soil-science map is finished the topographic elements remain unchanged. Due to human and financial resources which are needed for updating the geoscientific-maps in cycles comparing to ATKIS, it is impossible to offer a state of actuality which is a must, if the topography should be as actual as ATKIS.

3.4 DIFFERENCES

Between these data sets differences in acquisition mode, creation and updating occur.

The acquisition of topographic objects in ATKIS and the geoscientific maps is different. Different data sources are used in germany depending on the national mapping agencies. In Lower Saxony ATKIS is based on the German base map with a scale of 1:5000 (DGK 5) which offers a theoretical accuracy of 3 m (Harbeck, 1995). The topography in the geological map is based on the topographic map (TK 25) which is upscaled up to 1:10000 and copied to transparent foil to simplify the field-work and provide additional space for annotations (Ad-hoc, 2002).

Also the way of integrating updates differs. The topographic data in ATKIS is updated depending on the different object classes in very short terms. Objects like streets have top-actuality and are updated in periods between three and six months (LGN, 2003).

The geoscientific maps are handled and updated in a different way. Because of the thematic focus of both geoscientific maps the topography is just an element for giving orientation in the map and getting additional information which is needed for the geological or soil-scientific context.

The need of updating the thematic content in the geo-scientific data-sets is hardly comparable to the one of the ATKIS data-set, because the thematic content of the maps - without

considering the topography - will keep its actuality for decades or even longer.
To ensure the actuality of the topographic content in the geoscientific maps the update has to be carried out with the same periods like ATKIS. Parallel continuation will lead to semantic and geometric problems, e. g. misinterpretation or coordinate discrepancies due to different discretisation by different operators. Due to low change rates of the mere contents of the geoscientific maps, and also due to financial reasons, a periodical update can hardly be conducted. Therefore, an automatic adaptation of topographic updates from ATKIS to geoscientific maps is a very interesting possibility (Sester et al., 2003).

4. DATA INTEGRATION

Multiple acquisition of environmental objects lead to different major problems. Identical objects are acquired using various sensors and/or based on different data models results in heterogeneous data sets. Even multi-acquisition of data in a single data model may result in non fitting data sets due to different interpretation and discretisation (Walter, 1997).
The integration of datasets handled in a homogenous data-model can be fulfilled by geometric integration. The integration of inhomogeneous data-sets requires semantic integration first in order to avoid the comparison of "apples and oranges".

4.1 SEMANTIC INTEGRATION

Enabling the adaptation of updates from one data-set to another leads to the problem of integration of heterogenous data.
There are different types of data integration tasks (Walter & Fritsch 1996) :

- Type 1: integration of data-sets stemming from the same data-source with unequal updating periods,
- Type 2: integration of data-sets represented in the same data-model, but acquired by different operators,
- Type 3: integration of data-sets which are stored in similar, but not identical data-models,
- Type 4: integration of data-sets from heterogeneous sources which differ in data-modelling, scale, thematic content, …

The task of integrating ATKIS and geo-scientific maps can be categorized in type 4. Here, the matching also has to take into account, that the data can be modelled with different geometric primitives. For example all water-objects in the geoscientific data-sets like rivers and canals, are represented as polygons. The data-model of ATKIS offers two possibilities for representing these objects: below a specified width these objects are represented as lines with an attribute called BRG. This attribute shows the width of the water object in classes (e. g. BRG = "6" for a width of 3 to 6 meters), see Fig. 1. To ensure a correct and automatic adaptation of updates, the detection of changes and the relation between objects of different data-sets is a must. Therefore the common ground of the different integration tasks is that semantic integration has to be realised before the geometric integration can be processed, otherwise an integration procedure which is based exclusively on the geometry would compare polygons in a given area, which could be for example the representation of a forest on the one hand and a soil-classified ground on the other hand.
Due to these needs, a semantic catalogue has been created as a first step in which the thematic contents of the different data-sets are opposed allowing the definition of the common semantic content, e.g. objects which are represented in all of the data-sets are therefore matching candidates.

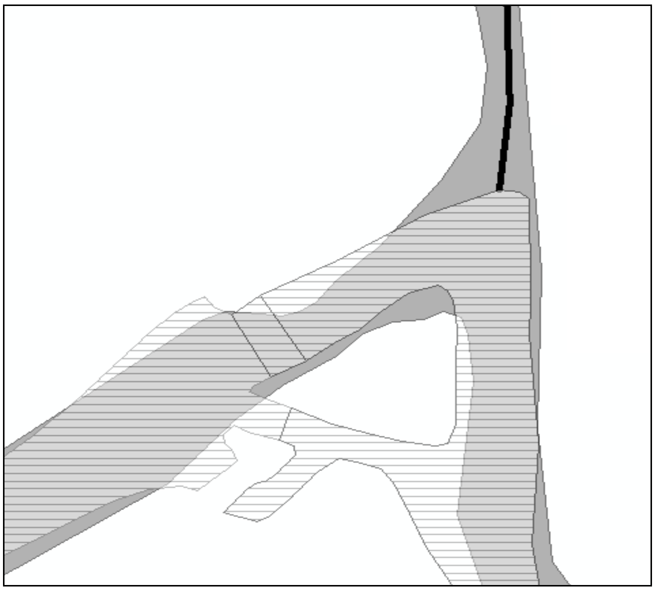

Fig. 1 : Different representations of water-areas in digital maps. River as line and polygon (dark line and hatched area) in ATKIS, and as polygon (solid fill) in the geoscientific-map.

A successful integration is based on a semantic model that compares the contents of a reference dataset (e.g. ATKIS) with the datasets in which the update should be maintained.
When – like in the given case – data-sets are compared that all have a topographic content, there are different object-classes which could be assumed to be suitable for building relations on: Streets, water areas, vegetation etc.

Objectclass	ATKIS (Objektart)	GK (Schicht)	BK (NRKART)
Polygon			
River	5101	qh/W////	-1
Canal	5102	qh/W////	-1
Lake/Sea	5112	qh/W////	-1
Line			
River	5101	1)	1)
Canal	5102	1)	1)
1) geological and soil-science map containing only polygons; rivers and canals are represented only when exceeding a certain width			

Tab. 1 : Semantic congruence between representations of water areas in ATKIS and geoscientific maps

In the first step of our investigation the object-class water is chosen as a candidate for matching between all data-sets. In later stages of this project other object classes will be examined whether they could serve as matching candidates or not. Water-objects have been selected, because they are present in all three data-sets. Therefore, this object class is used as a test bed for the first algorithms and methods.
The table 1 shows the attributes representing the water areas in the different data-sets. The geological attribute "qh/W////" is given in the standard description of soil-layers. These are separated into six values, e.g. stratigraphy (soil-age), lithography (main and additional description), the content of

humus-soil, the amount of lime and genesis, the way the soil layers were created.
The description from the soil-scientific map is called mapping unit. There are special values and terms for geo-scientific purposes. If a mapping unit has no soil-thematic label, but is of thematic or topographic interest, a special attribute is used for description. In the case of water areas this attribute is –1.

4.2 GEOMETRIC INTEGRATION

After the semantic integration has been finished, a group of semantic partners is delivered to be used as the first candidates for geometric integration in the next step. Depending on the underlying transformation between the two data sets, different matching methods can be applied: if the data sets are accurately geo-referenced, then a mere overlay and a check on the similarity of attributes will result in an identification of corresponding objects. Otherwise, also a (possibly unknown) transformation has to be determined first. Relational matching uses not only unary information (i.e. attributes), but also relations between objects, e.g. topological relation or quantitative relations like area-ratio. As the data sets are georeferenced, in our case the matching is realized by calculating the area-difference and centroid-distance between objects of geoscientific-maps and ATKIS.

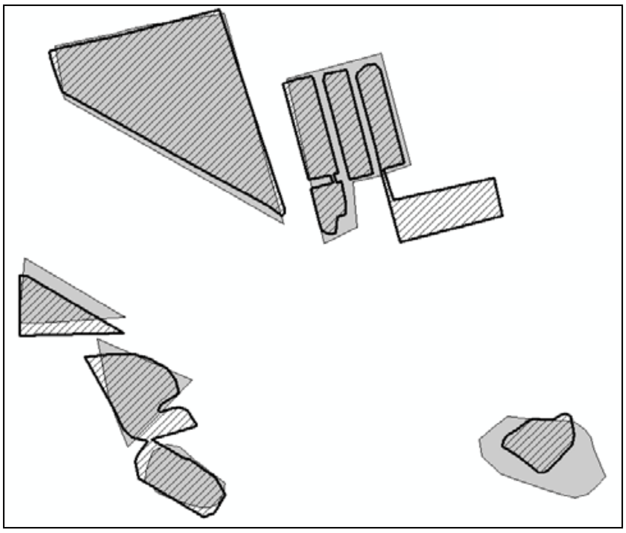

Fig. 2 : Superimposing water areas from ATKIS (dark border, hatched) and geological map GK 25 (solid fill grey)

Due to generalization effects, but also depending on the digitisation accuracy, objects which are represented as single objects in ATKIS, may correspond to several objects in the geoscientific map, and vice versa (cf. Fig. 2). These objects have to be matched by checking intersection-relations between objects.
In our case the search for partners in the different maps simply uses the distances between the middle-points of ATKIS-object and the candidates from the geoscientific map in order to select the first matching pairs. The pairs meeting the criteria of semantic and spatial conformity will be tested for the difference between their size.
Change detection through a mere intersection of corresponding objects and evaluation of resulting segments would most often lead to unsatisfying results as typically the objects do not perfectly overlap. Therefore, an object from the geo-scientifc data-base has to be adopted to the object from the ATKIS data-set using an algorithm based on a given transformation to achieve the best fit between both candidates. The iterative closest point (ICP) algorithm established by Besl & McKay (1992), which has been developed for fitting 3D-objects, is used here.

4.2.1 Iterative closest point method – ICP

Objects which have been selected through semantic and geometric integration and have been considered as a matching pair are fitted with the ICP algorithm. The objects from ATKIS are chosen as the reference object, because of the fact that they must not be changed and the level of actuality and accuracy is assumed to be higher than the one of the object selected from the geo-scientific map (which act as so-called fitters).
To avoid problems due to a large starting distance, the centroids of both objects are calculated and the geo-scientific object is 2D-translated (vertical and horizontal) to match the centroid of the reference object.
The implementation places a fixed number of points (railing points) on every segment on the polygon edges both reference and matching object. After creating a list of points on both objects these points are processed with the ICP algorithm. In the next step for every point on the fitter a point on the reference is searched, using the criterion of proximity.
These pairs are used as an input for a similarity transformation (Helmert-Transformation) achieving four new transformation parameters for translation, rotation and scale as result. The results are fed again into the process and the whole transformation is repeated iteratively.

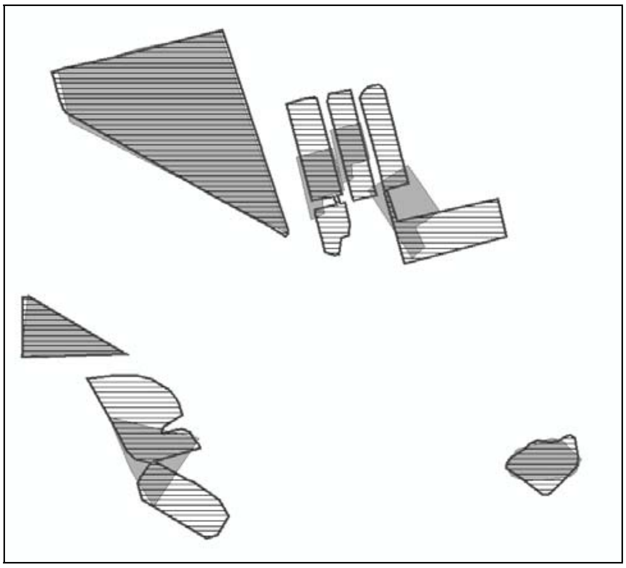

Fig. 3 : Object from geoscientific-map (BK) after transformation with ICP algorithm.

When no more variation in the transformation parameters occur the iteration stops and the fitter is aligned as best as can be to the reference object. At the end the distance of every railing-point on the fitter to the nearest segment of the reference-object is computed and a value for affinity is calculated.
Figure 3 shows the situation by adopting the ICP algorithm on the geoscientific data-set. Objects which are represented as single objects in both data-sets are fitted very good with ICP. Problems occur with unsolved 1:n or n:1 relations between the data sets. In the lower left corner two objects from the geo-scientific map have been matched to one ATKIS object, in the

upper right corner, one geoscientific object was copied three times and each copy has been fitted to the corresponding part in ATKIS. In the lower right, the object has been adapted considerably in size. Here, adequate thresholds are needed to determine when to accept a correspondence between the objects: the transformation parameters (e.g. scale not equal 1) give a good indication here.
This result shows the need of an algorithm which is capable of detecting 1:n and n:m relations between these data-sets, before the ICP algorithm is applied.

5. INTERSECTION AND EVALUATION

After matching and fitting selected candidates there has to be a change detection between the reference and the fitter. This is done by intersecting the objects, followed by an evaluation of the resulting segments. The following image (Fig. 4) shows the process.

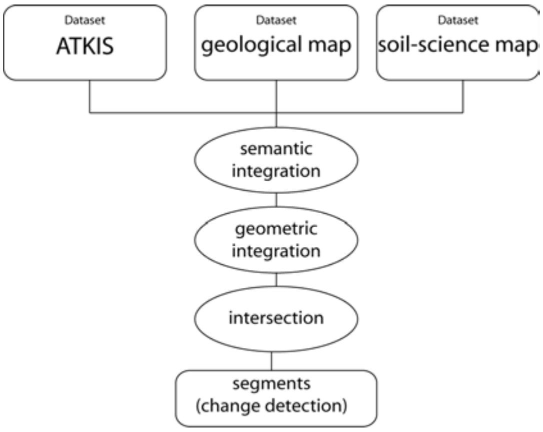

Fig. 4 : The process of integration and change detection

The segments can be categorized in three classes, according to the result of the intersection. Figure 5 shows an arroyo which has been straightened after the integration into the geoscientific map. Intersecting these objects would lead to three types of resulting polygons.

- Type I : Area is defined as water-area in both maps, no changes will be adopted,
- Type II : Area has been any type of soil, but is now defined as water-area, attribute of classification will be changed in geoscientific map,
- Type III : Area has been water-area, but is now updated, therefore a new classification is required.

While Type I. and II. require only geometric solutions to be updated in the geoscientific map, the situation with Type III. is slightly different. According to area-thresholds which have to be defined, the segment can be handled as geometric discrepancy resulting from inaccurate digitising, copying, or generalisation effects which occur during the production of the topographic map the object has been taken from.
Segments exceeding these thresholds must be handled in a different way according to an assumed semantic discrepancy which can not be corrected using only geometric operations. At this point geoscientific knowledge is mandatory.
Considering this special situation, an automatic process showing areas of type III discrepancies would be useful, to reduce the amount of human effort detecting the need of updates in certain areas of the geoscientific maps.
Automatic adaptation of change detection and integration is the goal of this approach. Therefore more semantic information has to be used to ensure a satisfying result from a full automatic process.

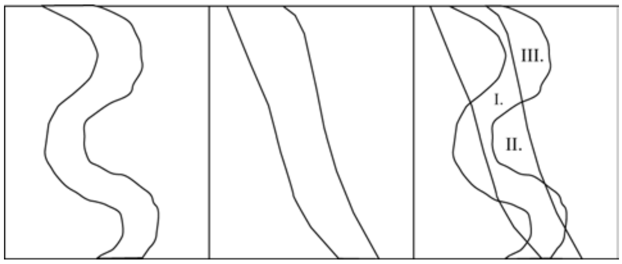

Fig. 5 : Intersection of an river, left: topographic situation in geoscientific map (time T1), middle: topography from ATKIS (time T2>T1), right: intersection of both objects

6. CONCLUSION AND OUTLOOK

In the paper the first approaches of semantic and geometric integration have been presented together with the implementation of the ICP algorithm for geometric adaptation. The first tests proved the object-class water to be a promising candidate for matching ATKIS, geological and soil-science map. With methods of attribute matching most of the related objects could be detected. The ICP-algorithm showed very promising results fitting matched objects and enhancing the results of the intersection process.
Ongoing research will expand the semantic model to cover all objects which are possible candidates for the semantic and geometric matching. Also, methods for the integration of linear or point objects have to be developed. Furthermore, the treatment of partial overlaps of corresponding objects has to be investigated (the general case of n:m-matchings). After the object-wise local adaptation of matching objects using ICP, the neighborhood of the objects has to be transformed accordingly. Also, investigations will be conducted to derive a sort of intermediate geometry of the two given objects, taking their accuracy into account. Up to now in our case, the ATKIS objects have been considered having higher quality, thus the geoscientific objects have been mapped onto them. Finally, a software prototype will be implemented which will be used as a test platform for testing different matching algorithms. More data-set will be processed and discussions with the external partners of this project will help to derive additional methods and parameters which are needed to create a fully-functional and automatic process.

ACKNOWLEDGEMENTS

This work is part of the GEOTECHNOLOGIEN project funded by the Federal Ministry for Education and Research (BMBF) and the German Research Council (DFG) under contract 03F0374A.

REFERENCES

Ad-hoc AG Boden, 1994. *Bodenkundliche Kartieranleitung*. Hannover, Germany, pp. 27-44.

Ad-hoc AG Geologie, 2002. *Geologische Kartieranleitung – Allgemeine Grundlagen.* Geologisches Jahrbuch G 9. Hannover, pp. 19 ff.

Badard, T., 1999. On the automatic retrieval of updates in geographic databases based on geographic data matching tools. In: *Proceedings of the 9th International Cartographic Conference*, Ottawa, ICA/ACI (Eds.), 1999, pp. 47-56.

Besl, P. and McKay, N.,1992. *A Method for Registration of 3-D Shapes,* Trans. PAMI, Vol. 14(2), pp. 239-256.

Brenner, C., 2000. Towards fully automatic generation of city models. In: *International Archives of Photogrammetry and Remote Sensing*, Volume 33, Part B3, Commission III, pp 85-92, Amsterdam, July 16-23, 2000.

Butenuth, M. & Heipke, C., 2003. Modelling the integration of hetergeneous vector data and aerial imagery. In: *Proceedings of ISPRS Commission IV Joint Workshop*, Stuttgart, Germany, September 8-9, 2003. Submitted.

Harbeck, R., 1995. Überblick über Konzeption, Aufbau und Datenangebot des Geoinformationssystems ATKIS. In: *(Kophstahl & Sellge 1995)*.

Kophstahl, E. & Sellge, H., Hrsg., 1995. *Das Geoinformationssystem ATKIS und seine Nutzung in Wirtschaft und Verwaltung – Vorträge anlässlich des 2. ADV-Symposiums ATKIS am 27. und 28. Juni 1995 in Hannover*, Niedersächsisches Landesvermessungsamt, Landesvermessung.

LGN, 2003. ATKIS in Niedersachsen und in Deutschland. In: *Materialien zur Fortbildungsveranstaltung Nr. 1/2003*, Hannover.

Mantel, D., 2002. *Konzeption eines Föderierungsdienstes für geographische Datenbanken.* Diploma thesis, unpublished, University of Hanover.

Sester, M., Hild, H. & Fritsch, D., 1998. Definition of Ground-Control Features for Image Registration using GIS-Data. In: *T. Schenk & A. Habib, eds, 'IAPRS', Vol. 32/3, ISPRS Commission III Symposium on Object Recognition and Scene Classification from Multispectral and Multisensor Pixels*, Columbus/Ohio, USA, pp. 537-543.

Sester, M., Butenuth, M., Gösseln, G. v., Heipke, C., Klopp, S., Lipeck, U., Mantel, D., 2003. New methods for semantic and geometric integration of geoscientific data sets with ATKIS – applied to geo-objects from geology and soil science. In: *Geotechnologien Science Report, Part 2*, Koordinierungsbüro Geotechnologien, Potsdam. Submitted.

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

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