# A MATCHING APPROACH FOR THE INTEGRATION, CHANGE DETECTION AND ADAPTATION OF HETEROGENEOUS VECTOR DATA SETS

**Guido von Goesseln**
Institute of Cartography and Geoinformatics
University of Hannover
Appelstr. 9a,30167 Hannover
Germany
guido.vongoesseln@ikg.uni-hannover.de

**Abstract**
Due to different acquisition methods and not synchronized updating periods the topographic content of geoscientific data sets and the German digital topographic map (ATKIS) differs in geometry, accuracy and actuality. In former times these differences between printed analogue maps were not as apparent as today, when different data sets are overlaid in modern GIS-applications. Integrating different data sets – in our case topographic and geoscientific data – allows for a consistent representation and for the propagation of updates from one data set to the other. To enable the integration of these data sets a workflow is in development, based on harmonization, geometric alignment, change detection, and updating. These steps are necessary to ensure consistency, but they are hardly practicable when performed manually.
Taking into account that corresponding objects from different data sets have been acquired at different points of time, it is obvious that parts of the geometry have been changed and that the comparison based only on the object geometry using standard algorithms will lead to unsatisfying results. The manual examination of the different data sets shows very obvious similarities in large parts of corresponding objects geometry. A simple matching using overlay for the selection of candidates is recorded as an XML relation set. According to these relations different geometric alignment strategies have been implemented and evaluated which allow the adaptation of the geometry. Based on different parameters and thresholds these strategies are capable of aligning geometries to avoid discrepancies, but they still enable the identification of changes which occurred between different acquisition steps. Beside the automated alignment and change detection, the automatic derivation of an updated geoscientific map is possible, which is performed using a rubber-sheeting transformation based on the different alignment strategies. This paper shows the integration workflow but concentrates on the different alignment methods.

## BACKGROUND

In former days many geoscientific questions have been answered using different analogue geoscientific maps. Using digitalization, these maps became digital data sets which can be integrated and compared using overlaying techniques in a GIS software. This offers new possibilities for solving geoscientific tasks. But it revealed other problems to the operator which did not meet the eye using the old analogue maps – geometric discrepancies. As it can be seen in Fig. 1 both data sets consist of homologue geometric objects, which are similar in shape and position, but the discrepancies are obvious. The reason for the geometric differences between the data sets used in this project, are based in the way of their creation. ATKIS is a digital topographic data set which is delivered from the german map agencies using different sources (e.g. aerial images, cadastral surveying) and is updated based on the different feature classes in periods from 3 months to one year.
As for ATKIS the topography is the main thematic focus, for the geoscientific maps it is either geology or soil science, these maps have been produced using the results of geological drills, according to these punctual informations, areal objects have been derived using interpolation methods based on geoscientific models. However they are related to the underlying topography. The connection between the data sets has been achieved by copying the thematic information from topography to the geo-scientific maps at that point of time the geological or soil-science information is collected. This is done by using up scaled copies (1:25.000 to 1:10.000) of topographic maps. The selection and integration of objects from one data set to another one has been performed manual and in most of the cases the objects have been generalized by the geoscientist. While the geological content of these data sets will keep its actuality for decades, the topographic information in these maps do not: In general, topographic updates are not integrated unless new geological information has to be inserted in these data sets. These different ways of acquisition and updating lead to geometric discrepancies, which make it very hard and sometimes impossible to use them in combination with other geoscientific data sets from different domains or with digital topographic maps delivered from authorities. In a project of the German Ministry of Education and Research under the headline “GEOTECHNOLOGIEN”, a research group at the University of Hannover, consisting of three institutes from surveying and computer science, is dealing with the problem of data integration, applied to data sets from topography, geology and soil science. The project tackles with different aspects of

XXII International Cartographic Conference (ICC 2005) A Coruña, Spain, 11-16 July 2005
Hosted by:
The International Cartographic Association (ICA –ACI)
ISBN: 0-958-46093-0 Produced by Global Congresos

data integration, namely integration of different vector data sets, integration of vector and raster data, as well as providing an underlying data structure in terms of a federated data base, allowing a separate, autonomous storage of the data, which are linked and integrated by adapted reconciliation functions for analysis and queries on the different data sets (Sester et al., 2003).

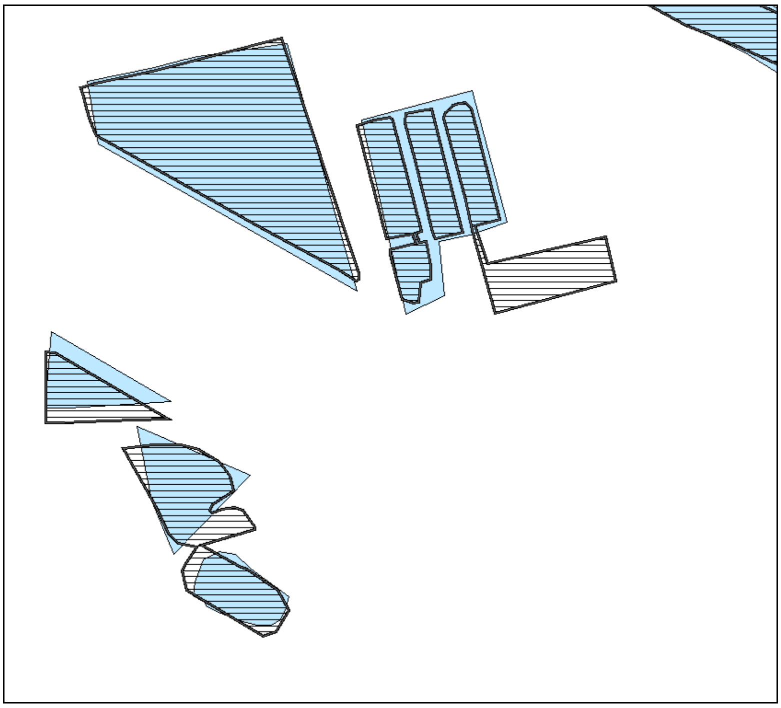

Fig. 1 : Simple superimposition of ATKIS (dark border, hatched) and geological map GK 25 (solid fill).

Experts from different geoscientific domains in Germany take advantage of the geological (GK) and the soil-science map (BK). These maps have a very strong thematic focus, but they do not contain the amount of topographic content, which is mandatory for different tasks to be solved. Therefore these data sets are combined with the german digital topographic data set (ATKIS). Unfortunately the geoscientific data sets have been digitized from analogue maps and they differ in acquisition time, representation type and temporal consistency which makes integration hardly possible. Therefore certain tasks have been defined:

- Data integration with the consideration of semantic content,
- Removal of geometric discrepancies based on map creation processes using alignment strategies,
- Automated detection and evaluation of changes in the underlying topographic content in the geoscientific maps,
- Generation of an updated geoscientific map.

The geometric discrepancies can be explained as changes which occurred during the different points of time the data sets have been acquired, or as inaccuracies which are reasoned in the way of the data set creation. The degree of discrepancy can be used to explain the differences and select appropriate methods for alignment. The appropriate method must not align every discrepancy which occurs in the data set, by evaluating the alignment results it must be possible to select between discrepancies based on acquisition or changes which occurred to the real world object. Otherwise the detection of changes which have to be inserted to the map as new geoscientific content would be blurred by the alignment process.

## RELATED WORK

Data can be integrated and fused for mutual benefit: Walter & Fritsch, (1999) present an approach that fuses two different data sets with road information with the aim of mutually exchanging attributes of the two data sets. The integration of vector data and raster data is being investigated in a GEOTECHNOLOGIEN partner project with the aim of enriching a 2D-vector data set with 3D-information (Butenuth & Heipke, 2003).
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). A conceptual framework for the integration of geographic data sets, based on a domain ontology and surveying rules, was developed for update propagation between topographic data sets (Uitermark, 2001). While this

approach is primarily intended to work with a matching approach based on the geometry, there are concepts which also take relation and data attributes into account (Cobb et al., 1998). There are modern approaches using expert expressions to integrate time-series where classification has changed (Comber et al., 1994). The alignment of geometries in this framework is performed using different approaches. For geometries which can be aligned using a four parameter transformation the iterative closest point algorithm (ICP) has been implemented, which originally has been developed by (Besl & McKay, 1992) for aligning 3D objects. In Kohonen (1997) the learning vector quantization (LVQ) approach has been presented which associates an entire feature vector sequence, instead of a single feature vector. The usage of geometric harmonization for cadastre data sets has been presented by Hettwer & Benning (2000). The adaptation of maps using a rubber-sheeting algorithm was presented in Doythser (2000).

## APPLICATION FRAMEWORK

It has been shown by Yuan & Tao (1999) that a conflation strategy based on modules to solve special integration problems seems very promising. The module-based design also offers the possibility to use the results developed for this project for other vector-vector conflation tasks.

At the beginning of the integration process the semantic content of all data sets has been compared. The topographic elements "water-objects" which are represented in all of the three data sets are selected and will be used as candidates for the matching process. This selection is mandatory to avoid comparing “apples and oranges” and has to be the first step to ensure a successful integration.

An area-based matching process is used for the creation of links between object candidates. These links are stored in a database using a XML-schema, following by an alignment process which will reduce geometric discrepancies to a minimum to ensure satisfying results in the subsequent intersection process. A rule-based evaluation of the intersection results is used for change detection.

## INTEGRATION

Water objects in ATKIS are represented in two different ways: Water areas and rivers exceeding a certain width are represented as polygons. Thinner rivers are digitised as lines with additional attributes, referring to some classified ranges of widths. The representation of water objects in the geo-scientific maps is always a polygon. These differences are solved using harmonization strategies.

Another problem is the representation of grouped objects in different maps. For a group of water objects, e.g. a group of ponds, the representation in the different data sets could either be a group of objects with the same or a different number of objects, or even a single generalised object (see Fig. 2). Finally, also objects can be present in one data set and not represented in the other. All these considerations lead to the following relation cardinalities that have to be integrated: 1:0, 1:1, 1:n, and n:m. After the corresponding relations have been identified, each selection set will be aggregated, so they can be handled as 1:1 relations (Gösseln & Sester, 2004).

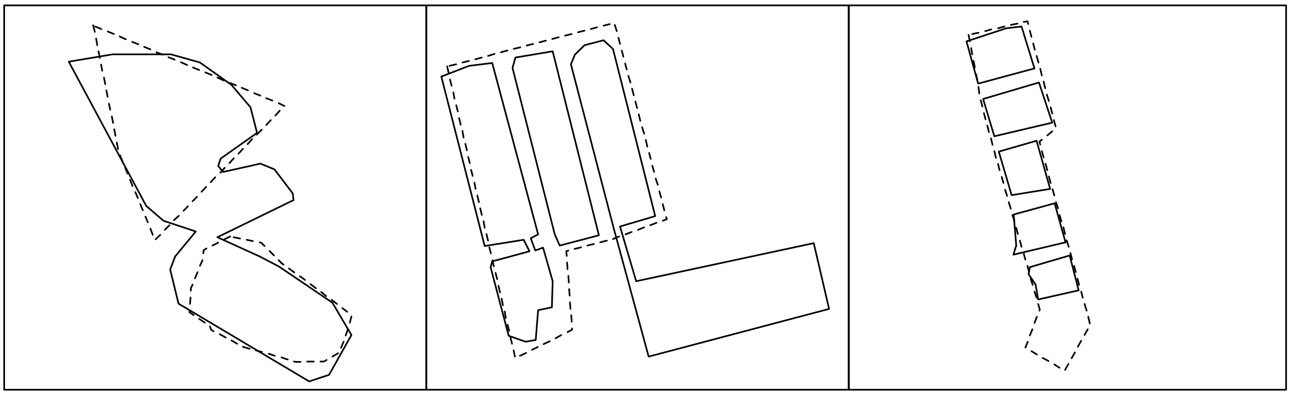

Fig. 2 : Different representations - ATKIS (solid line), GK (dotted line)

The automatic investigated links can be visualized to the operator, enabling a manual correction of the derived links. The object links which are the result of an automated process can be evaluated using a GUI based application. With this software every relation set can be inspected and edited, if the automated process failed to build up the suitable correspondences between the selected data sets.

Due to the fact that the objects from all three data sets are representations of the same real world objects, they show apparent resemblance in shape and position.

Nevertheless the alignment of the geometries is required after the evaluation of the matching results. As it will be described later, there is no geometric alignment method which is capable of covering all alignment tasks. Therefore the alignment technique offering the most suitable result can be selected for every single relation set.

## GEOMETRIC ADAPTATION

Objects which have been considered as a matching pair could be investigated for change detection using intersection strategies. At this stage the mentioned differences will produce more problems which are visible as discrepancies in position, scale and shape. These discrepancies will lead to unsatisfying results and make the evaluation of the resulting elements almost impossible, this will evoke an immoderate estimation of the area investigated as change of topographic content.
Therefore a geometric adaptation will be applied, leading to a better geometric correspondence of the objects. For these adaptation processes thresholds are required which allow the reduction of discrepancies which are based on map creation, but will not cover the changes which happened to real world objects between the different times of data acquisition.

### Iterative closest point (ICP)

The iterative closest point algorithm (ICP) developed by (Besl & McKay, 1992) has been implemented to achieve the best fitting between the objects from ATKIS and the geo-scientific elements using a rigid transformation.
The ICP algorithm has been developed to align three-dimensional objects using a 7 parameter transformation. In this case the problem is reduced to a 2D problem which in our case requires four parameters (position, scale, orientation). The implementation places points (railing points) with a fixed distance on the contours of the object to be aligned. For every railing point the closest point on the corresponding object is selected. These coordinate pairs are taken as an input value for a iterative similarity transformation (Helmert-transformation) leading to four new parameters as result. These transformation parameters are evaluated after each calculation; the iteration stops when no more variations in the four parameters occur.
At the end of the process the best fit between the objects using the given transformation is achieved, and a link between corresponding objects in the different data set is established. Evaluating the transformation parameters allows to classify and characterize the quality of the matching: in the ideal case, the scale parameter should be close to 1; also rotation and translation should be close to 0. Assuming, that the registration of the data sets is good, these four parameters exactly meet the reasons for the geometric discrepancies and differences between ATKIS and the geoscientific maps, which are reasoned by the manual creation of the maps. Therefore a greater scale factor can be an indicator, that the difference between two objects is not based on map creation, but on a change at the real world object, which occurred between the data acquisition of the geoscientific and the ATKIS data set (Goesseln & Sester, 2004).
At the end of the process the best fit between the objects using the given transformation is achieved, and a link between corresponding objects in the different data set is established.
The application of the iterative adaptation using the ICP approach showed very good results and revealed the possibility of reducing the amount of objects which have to be evaluated manually. However there are some situations where this approach does not generate sufficient results.

### Dual interval alignment (DIA)

Objects like streams or rivers which are represented in the data sets as large groups of connected polygons can be aligned using the ICP approach, but the result will not be satisfying and serious geometric discrepancies will occur at objects which cover several map-sheets or at least touch the map boundaries.
Therefore a different approach, called DUAL INTERVAL ALIGNMENT (DIA) has been implemented, aligning the geometry of matched features calculating the transformation of single vertices to enable the alignment of local discrepancies. Corresponding objects which have been assigned as representations of the same real world object through the matching process are investigated based on their vertices. For every point in one object the nearest neighbour in the corresponding partner object is determined using the criterion of proximity. The conformation approach evaluates the distance between these coordinates. Based on an interval which is predetermined by the human operator according to map-scale and the threshold which defines the largest change in geometry which will be suitable for the geoscientific data set. As it can be seen in Fig. 3, for each point ($P_C$) from object C and the corresponding point ($P_R$) of the linked object R, the point transformation is calculated based on the euclidean distance (d) between these points. The new coordinates are determined taking interval ranges a and b into account. Points within the first distance interval ($0<d<a$) are aligned to a single point, a distance falling into the second interval ($a<d<b$) will lead to an approximation of the selected points. Points with a distance beyond b will not be adapted (see Eq. 1).

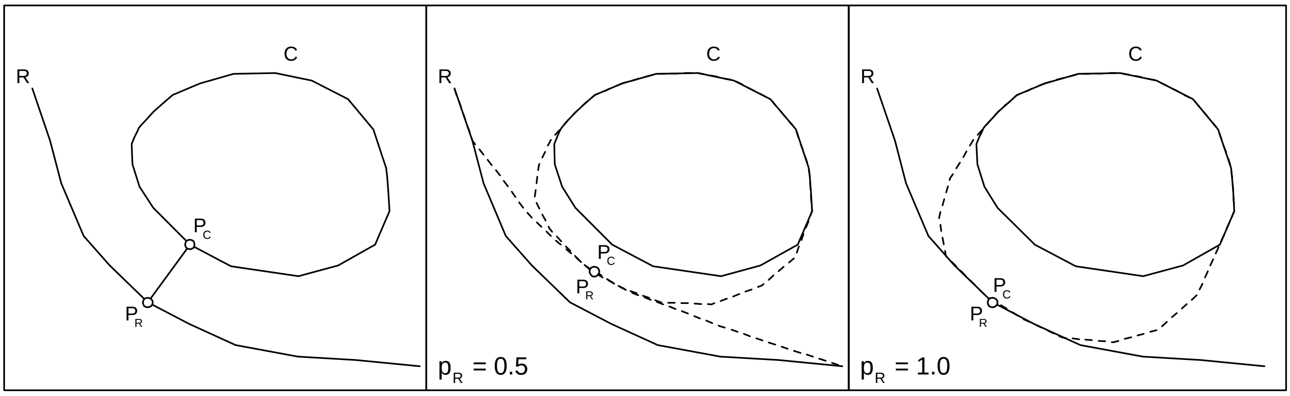

Fig. 3 : Application of DIA for the partial alignment of object geometries (schematic).

The integration of a weight to the alignment process does not only take different accuracies of geometries into account, but opens this approach to a much wider range of conflation tasks. E.g. in this project one data set is handled as a reference data set and must not be changed. This can be achieved by setting weight p to 1. In other cases, when two data sets have to be aligned and no single data set can be selected as reference the alignment is performed using the common idea of conflation by aligning two data sets to a new improved data set.

$$(x,y)_{P_{new}} = \begin{cases} (x,y)_{P_{old}} + \left(\pm\Delta(x,y)\cdot p\right) & ,0 < d < a. \\ (x,y)_{P_{old}} + \left(\pm\Delta(x,y)\cdot p\right)\cdot f(d) & ,a < d < b. \\ (x,y)_{P_{old}} & ,b < d. \end{cases}$$

$$f(d) = \left(\frac{b-d}{b-a}\right), \quad p(P_c) = 1 - p(P_r), \quad \Delta_{RC}(x,y) = (x,y)_R - (x,y)_C$$

Eq. 1 : Calculation of point transformations using DIA

## Segment orientation

Calculating the shift distance based on the nearest neighbour enables very good alignment result, but can result in topologic errors. As it can be seen in Fig. 4 a relational error occurs between two rivers from different data sets representing the same real world object. By using the conformation approach and evaluating the distance with the described interval, the distance criteria fails by aligning the right side of the stream to the left side of its corresponding partner. This requires the integration of an additional criteria. In a first step the orientation of two corresponding polygons will be compared. In the following step, the orientation of the polygon segments will be calculated for all corresponding objects.

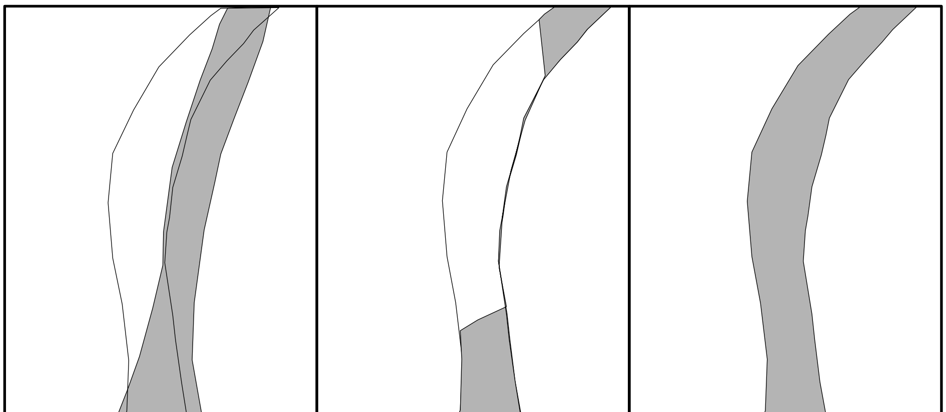
Fig. 4 : Left - ATKIS (dark border) and GK (filled) river, Middle - GK object aligned, Right - GK object aligned with estimating segment orientation

If a point and its corresponding partner are selected using the distance criterion, the direction to each corresponding successor will be calculated. If the difference between these directions exceeds a given threshold, the points must not be aligned due to the assumption that they do not represent the same "side" of the real world object.
As it can be seen in Fig. 5 the points An and Qn would be appropriate alignment candidates regarding the distance criteria, but by evaluating the segment orientation, An and Pn will be selected as correct candidates. To ensure the result for objects which are modified using the DIA approach, the digitization orientation will be identified in advance.
Comparing the adaptation approaches ICP and DIA, each is suitable for a different kind of objects in this project. ICP matches the idea, that the majority of the geometric discrepancies is caused in the way the geoscientific data sets have been created by integrating topographic elements through manual copying. The resulting parameters can be used for the investigation and the evaluation of the influences which were responsible for the geometric discrepancies:
An object which can be aligned by just using translations with a small change in scale can be judged as minor error based on manual copying. A larger scale factor can reveal topographic changes on the real world objects, which have to be investigated by a human operator.

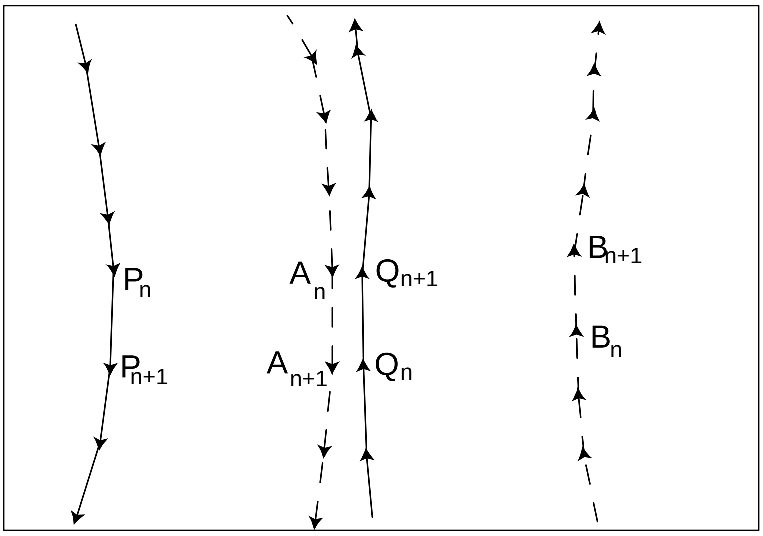

Fig. 5 : Consideration of the direction between point and successor to enhance the alignment process. Segment of two river streams from ATKIS (solid line, P&Q) and GK (dotted line, A&B).

The transformation, the implemented ICP algorithm is based on, does not give satisfying results for larger, irregular shaped objects like rivers, boundary of soil classes, etc. or objects, that match only partially. The combination of both approaches delivered very good results, offering the possibility to assess the geometric discrepancies by evaluating the resulting ICP-parameters, and aligning large object groups or partially matching objects using DIA.

### Rubber sheeting

After the corresponding objects have been adapted individually, also the neighbourhood of these objects have to be adapted as well. For that reason the point-transformations which have been applied to the vertices of the object geometries will be used as shifting-vectors for the rubber-sheeting process. For each vertex in the origin geoscientific data set which has not been adapted so far, the new position is calculated based on these shifting-vectors, weighted by the distance between the start-point of the vector and the position of the vertex to be translated. This process in the framework ensures the consistent presentation of the geoscientific data set.

## ALIGNMENT RESULTS

In Fig. 6 the results of the different alignment methods can be seen. The ICP algorithm using an iterative four-parameter transformation is very suitable for the alignment of objects which already have a similar geometry. The alignment parameters which are the results of the ICP algorithm can give a first hint whether the geometric discrepancies are due to map creation and acquisition methods (a., d.) or to changes which occurred to the real world object (c.). The resulting scale factor which was calculated for the alignment of object c. was rated as too large and therefore no alignment was performed. Of course changes of the topography can not be discovered by simple evaluation of these parameters. For object b. the algorithm achieved a best fit with four parameters below certain thresholds, but the remaining differences between the geometries still have to be corrected.

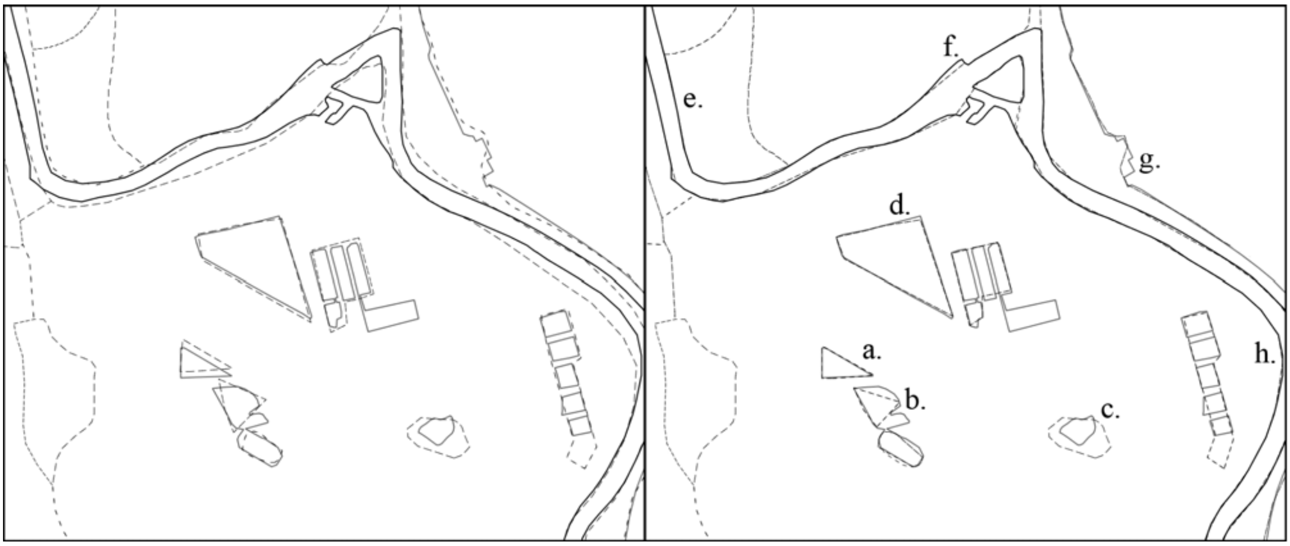

Fig. 6 : Result of the approach, GK 25 (thin, dotted line) aligned on the reference German digital topographic map (ATKIS, dark lines).

The DIA implementation showed very good results for local discrepancies which can not be corrected using the four-parameter-ICP which aims for the best alignment of the whole object. There is no single four-parameter transformation which is capable of adjusting one GK object to its corresponding partner in ATKIS so that parts e,f and h would be properly aligned. Due to the fact that the DIA algorithm in the current version is only working on existing vertices without inserting additional ones, there are some areas (g) where an appropriate alignment is not achievable.
After the alignment process has been completed an intersection of corresponding objects is used for the change detection. Without the alignment process in advance the mentioned discrepancies would lead to unsatisfying results and make the evaluation of the resulting elements almost impossible. The intersection is performed on all types of topographic elements which are represented in the data sets. The results of the intersection process will be evaluated and according to their semantic attribution sorted into three different classes.
It is expected, that a high degree of automation can be achieved with this process. In some situations there will be an automatically generated suggestion from the algorithm, however the expertise of a human operator will still be mandatory in some cases in order to commit or propose another solution (Gösseln & Sester, 2004).
All discrepancies which are still existent after the alignment process will be interpreted in the following step which is performed using an intersection approach and an automated evaluation which is capable of providing solution proposals to the operator (Goesseln & Sester, 2004).

## CONCLUSION

The implementation of the workflow in a software protoype using JAVA and an open source software GIS (JUMP, 2005) will be used as a test bed for the evaluation of the implemented algorithms, the derivation of suitable parameters and thresholds and it will be used as a prototype which can be used by external project partners.
The implementation of the geometric comparison and the derivation of object links, together with the ICP and DIC alignment followed by rubber-sheeting and the evaluation process showed already very good results (see Fig. 6).
At this point of the project one data set is selected as reference data set, which remains unchanged while the candidate data set is adjusted, but this approach can also be adopted to other vector based conflation tasks requiring an intermediate geometry. This process will lead to a local adaptation of the individual corresponding objects, but also of their local environment. Too large discrepancies of the shape boundaries will be considered as outliers and can be treated in the subsequent overlay and analysis step.
The combination of diverse algorithms enables the aligning of different kind of object-geometries. While ICP is very suitable for objects with similar geometries and the resulting parameters can be used as a first indicator for the change detection, DIA enables the partial alignment of large objects and aggregated groups. The need for partial alignment also occurs at object boundaries: e.g. a geoscientific object ends at a river or road. This means, that these features have a part of the river or road boundary in common. Due to the fact that only the geometry of linked objects is changed, the integration of a rubber-sheeting algorithm is used to adjust the neighbourhood. The module-based workflow offers the possibility of adopting the workflow to different situations, the integration of the geoscientific data sets to the german digital map as a reference and the alignment between different geoscientific data sets with the need of an intermediate geometry. This project is performed in close relationship with external partners from geoscientific domains which require an automated process for change detection and map adaptation, because a manual process would exceed the

available human and financial resources, and therefore only an automated workflow is suitable for detecting changes in topographic content and create an updated data set.

## ACKNOWLEDGEMENTS

This is publication no. GEOTECH-133 of the GEOTECHNOLOGIEN project funded by the Federal Ministry for Education and Research (BMBF) and the German Research Council (DFG) under contract 03F0374A.

## REFERENCES

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

Butenuth, M. & Heipke, C. 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.

Cobb, M., Chung, M.J., Miller, V., Foley, H.I., Petry, F.E., and Shaw, K.B., 1998. A Rule-Based Approach for the Conflation of Attributed Vector Data, GeoInformatica, 2(1): p. 7-35, 1998.

Comber, A., Fisher, P. & Wadsworth, R., 2004. Integrating land-cover data with different ontologies: identifying change from inconsistency. Int. J. Geographical Information Science, 18(7), p. 691-708, 2004.

Doytsher, Y. 2000. A rubber sheeting algorithm for non-rectangular maps, Computer & Geosciences, 26(9-10), p. 1001-1010, 2000.

Kohonen, T., 1997. Self-Organizing Maps. Springer, 1997.

Gösseln, G. v. & Sester, M.: Integration of geoscientific data sets and the german digital map using a matching approach, In: International Archives of Photogrammetry and Remote Sensing, Vol. 35, ISPRS, Istanbul, 2004.

Hettwer, J. & Benning, W. Nachbarschaftstreue Koordinatenberechnung in der Kartenhomogenisierung, Allg. Verm. Nachr. 107:194-197, 2000.

JUMP, 2005. The unified mapping platform, http://www.vividsolutions.com, (visited April 2005).

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., Goesseln, 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 No. 2, Hrsg. Koordinierungsbüro Geotechnologien, Potsdam, 2003.

Uitermark, H., 2001. Ontology-based geographic data set integration. Doctor thesis, Deventer, Netherlands.

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

Yuan, S. & Tao, C. 1999. Development of conflation components. In: Li, B. et al., eds., Geoinformatics and Socioinformatics, The proceedings of Geoinformatics´99 conference, Ann Arbor, pp. 1-13.xx