Concepts for quality assurance during mobile online data acquisition

Stephan Mäs, Wolfgang Reinhardt, Admire Kandawasvika, Fei Wang
AGIS - Arbeitsgemeinschaft GIS
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
D-85577 Neubiberg, Germany
http://www.agis.unibw-muenchen.de
Stephan.Mäs@unibw-muenchen.de
Phone: +49-(0)89-6004 2182, Fax: +49-(0)89-6004 3906
Wolfgang.Reinhardt@unibw-muenchen.de
Phone: +49-(0)89-6004 2450, Fax: +49-(0)89-6004 3906
Admire.Kandawasvika@unibw-muenchen.de
Phone: +49-(0)89-6004 3873, Fax: +49-(0)89-6004 3906
Fei.Wang@unibw-muenchen.de
Phone: +49-(0)89-6004 2716, Fax: +49-(0)89-6004 3906

SUMMARY

This paper describes the conceptual aspects and the development of a sophisticated, open standards-based mobile geodata acquisition system. The system offers services supporting field workers from different geoapplication domains for example geology, geophysics, utilities or others. One notable aim of the system is interoperable access to heterogeneous databases in the field. Therefore a generic acquisition concept, which flexibly adjusts to different application schemas and data models, has been developed in a prototype. The basic idea of the concept is to download schema information and to adjust the acquisition process to this information at runtime. The employment of such a concept makes investigations on the quality assurance during mobile data acquisition necessary. Since the schemas normally describe what kind of data has to be collected and how this data has to be organized, they leave out information about specific integrity rules and constraints contained in the data model. Therefore other methods to formalize such restrictions are required. As a possible way to describe integrity constraints in a formal way ontologies are introduced. It is pointed out how such constraints can be defined in OWL (Web Ontology Language), respectively in its proposed extension OWLR (OWL Rules Language), and how these rules can be used as an enhancement to the existing data schema information. Finally, the advantages of the proposed concepts are illustrated for a landslide application scenario.

This research work is being carried out as part of a bundle research project called "Advancement of Geoservices" funded by the German Ministry of Education and Research (BMBF-Bundesministerium fuer Bildung und Forschung) and German Research Foundation (DFG-Deutsche Forschungsgemeinschaft). The project consortium comprises University of Osnabrueck (Research Centre for Geoinformation and Remote Sensing), University Karlsruhe (Institute for Photogrammetry and Remote Sensing), EML (European Media Laboratory) Heidelberg and University of the Bundeswehr Munich (AGIS GIS Lab).

KEYWORDS: *LBS, mobile data acquisition, quality, integrity constraint, ontology, OWL*

INTRODUCTION

Mobile acquisition of geodata usually requires experienced users with knowledge about existing data and the underlying data models. All necessary data has to be examined and synchronized before going to the field. Later on, if any unexpected circumstances occur, there is no possibility for further download of data more relevant to the situation at hand. After finishing the acquisition, incorrectness and incompleteness of data are often only recognized when the fieldworker is back in the office, during the server update. In case, the newly captured data does not meet all the conditions of the data model and corresponding integrity constraints, the worker might even have to visit the exploration site again. This situation is quite typical for current acquisition systems.

Within the project "Advancement of Geoservices", the application of current evolutions in information and communication technologies and their usability for mobile data acquisition are investigated (Breunig et al. 2003 + 2004). Existing Location Based Services (LBS), e.g. presented in (Sayda et al. 2002), make use of these developments and offer information adjusted to the user's position and environment. The paper is presenting a field service extending the functionality of such LBS by methodologies for in-field accessing, analyzing, acquiring, quality checking and online updating of geospatial data. It is shown, how a continuous workflow from position determination and object acquisition to the transaction of the newly captured data to the databases can be realized.

Main issues of mobile data acquisition

One of the main objectives within the project "Advancements of Geoservices" is the development of concepts for the online mobile acquisition of geodata. In particular, the following aspects are taken into consideration:

- Establishment of an online access to all relevant databases from the field site
- Definition of an open platform based on standards, thus avoidance of proprietary developments and interfaces
- Design of an architecture for a mobile client
- Development of a generic acquisition concept, which flexibly adjusts to different application domains and data models
- Definition of main workflows like initial acquisition or update of data
- Extensive and throughout quality assurance for these workflows

In general, the proposed mobile acquisition system should enable mobile users to efficiently carry out their field work. The system should allow for:

- optimal, location-based access to the geospatial databases residing in a single or distributed servers,
- acquisition of new data by observations or measurements (i.e. geometry and related attributes for newly captured features/objects),
- automatic online update of the databases while in the field and still being consistent with the requirements imposed by the business or organization's data model or schema,
- and analysing and checking of the overall quality of the newly acquired data in the field.

The paper is focusing on quality assurance aspects for the mobile data acquisition process. Therefore the following two subsections will give a short introduction to the system architecture and the generic acquisition concept of the mobile data acquisition system developed in the project. Based on that, quality assurance concepts for the introduced system are discussed. Finally, the practical advantage of the quality assurance approach is demonstrated by giving some examples of integrity constraints within a landslide application scenario.

SYSTEM ARCHITECTURE.

The overall architecture of the system is presented in *Figure 1*. The selected conventional, standardized GIS client /server interfaces like WMS and WFS (OGC 2002a) can also be applied for mobile services. On basis of the available mobile communication technologies like WLAN, GPRS, UMTS and Bluetooth, it is possible to network different mobile system components. It's conceivable that the bandwidth of UMTS and WLAN supports the transfer of larger amounts of data. So the problem of the small bandwidth faced by previous technologies is solved and the principal requirements for online mobile access to heterogeneous databases are meet. The usage of standardized interfaces and therewith the avoidance of proprietary developments leads to an open structure of the GIS platform (Plan et al. 2004).

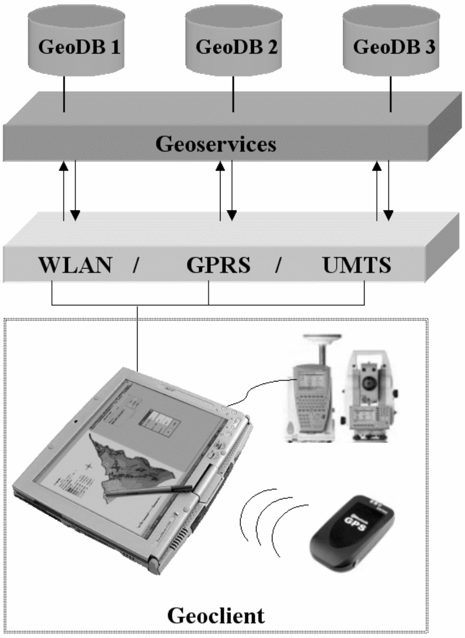


Figure 1: Architecture of the mobile acquisition system

The client equipment should be portable, enable visualization under outdoor conditions and have convenient possibilities for data input. Rugged devices like tablet PCs or laptops with pivoting touch screens fulfill these requirements and offer the necessary performance for the application aimed at. Often communication components like WLAN or Bluetooth are already included in the standard equipment. Since the collection of new data will be done using different sensors, another main issue of the implementation design of the system architecture is the interoperable interfacing to these various instruments. Therefore research is also focusing on the interoperable handling or usage of different types of geodetic sensors like GPS, total stations, laser scanners, extensometers, etc..(Kandawasvika & Reinhardt, 2005)

GENERIC ACQUISITION CONCEPT

The architecture of the system allows for applications in different spatial domains and enables the fieldworker to access any information source that might be of interest for the current use. But the possibility to have interoperable access to heterogeneous databases and not being restricted to a certain, well-known data model has great influence on the whole client application structure and the acquisition process. At first, it requires the client to download schema information at runtime. Such schema may contain all necessary details about the modeled objects/features, their geometry and attributes as well as the relations between objects of one or different types. Moreover, the client application has to be able to adjust to the requirements imposed by the data model. In particular, the measuring process and the templates for input of further attributes must be flexible and adaptable. *Figure 2* schematizes how the proposed application solves these problems.

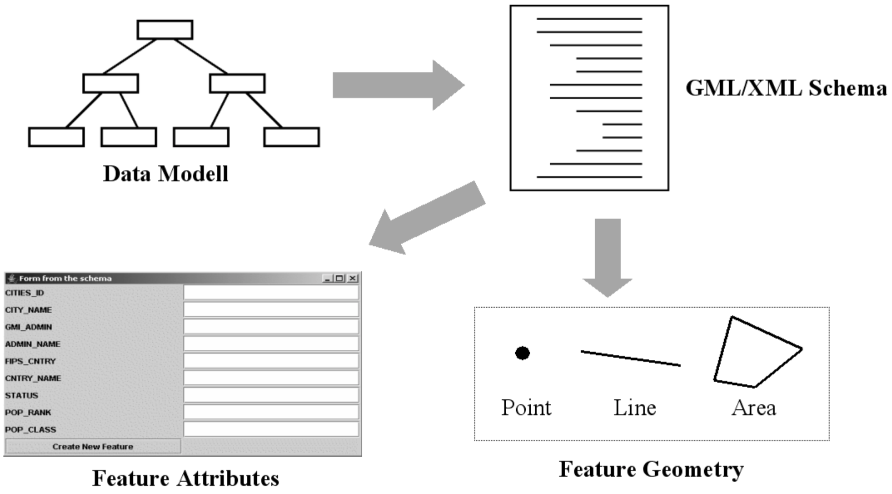


Figure 2: Attribute and geometry extraction for generic acquisition

The server connection and data flow are based on OGC standards like WFS and GML (OGC 2002a & 2002b). For the acquisition of an object, the user firstly has to select the desired object class, i.e. feature type. Information about the classes the server is supporting and some other basic information is available through the WFS capabilities document. The particular data schema (i.e. schema information of the selected feature type) is specified in XML-schema documents. Such XML-schema is used to describe what kind of data has to be collected and how this data has to be organized. The client application can download these documents for the defined feature types using the WFS interface. With the information contained in the XML-schema, it is possible for the client application to adjust the acquisition process in regard to the required attributes, geometry types and relationships of a particular feature type and to guide the user through the whole data collection procedure. The templates for the input of attribute values are generated automatically at runtime and the process of measuring geometry elements is adjusted to the requirements of the feature type currently being measured. Comments, for example on the meaning of certain attributes (i.e. their semantic), can be included in the XML-schema and requested by the user during the acquisition process.

Through the exploitation of generic data schemas, the system is flexible and independent of specific user domains. The decision, which data stock is most adequate to the current situation and requirements, can be made spontaneous in the field. The addressed WFS interface servers are self-describing and the acquisition process can flexibly be adjusted to the particular data model.

QUALITY ASSURANCE CONCEPTS

The mobile interoperable access to heterogeneous geodata bases and their update from the field has far reaching consequences for the data acquisition process. As mentioned before, this approach provides the possibility to check the newly recorded data in terms of quality and reliability directly in the field, which makes quality management investigations necessary. Useful automated methods are required to assure the quality of the data stock during online data recording. Every transaction of newly acquired or updated objects has to be checked against well-known quality parameters like consistency, completeness, correctness and accuracy. Although an overall quality management is of interest in this project, we will restrict ourselves in this paper on the definition and consideration of integrity constraints in quality checks, in particular as extension to the usual XML-schema validation. This validation of the captured data against the XML-schema will only assure that:

- every object has the right geometry type (according to the definitions specified in the data model),
- all necessary attributes and relations are considered and
- the attribute values conform to the defined attributes data types.

Definition of integrity constraints

It is obvious that, by far, not all quality aspects are considered in this list. For example, illegal topological relations like overlapping of the area objects lake and vegetation will not be recognized through a validation against the XML-schema downloaded from a WFS. A definition for such topological integrity constraints, which refer to the semantic of the object classes, has been given by (Ubdea & Egenhofer, 1997). They defined a topological constraint as an association of two geographical object classes, a relation, which is in this particular case a topological relation, and a specification (See *Figure 3*), which can have one of the following values:

1. Forbidden (i.e. zero times allowed)
2. At least n times
3. At most n times
4. Exactly n times.

Constraint = (Entity class1 , Relation , Entity class2 , Specification)

Figure 3: Definition of integrity constraints (Ubdea & Egenhofer, 1997)

In literature, there are many examples of the application of these or similar topological constraints for consistency assessment of databases. E.g. (Servigne et al. 2000) and (Ubdea & Egenhofer, 1997) worked on automatically identifying and correcting topological errors within existing data. (Cockcroft 2004) focused on the definition of constraints and consistency checking therewith during the data entry. She extended the former constraint definition by adding further attribute constrictions to address only particular objects, like for example “a pipe > 14 cm in diameter”. In (Pundt 2002) it is illustrated, how semantic rules can be used for the simultaneously control of the data collection procedure during fieldwork, e.g. for resolving conflicting attribute values of an object. With such plausibility checks it is possible to warn the field worker immediately, if any semantic inaccuracies occur. Pundt suggested to include the description of integrity constraints directly within the classes, e.g. through class paradigm like RDF (Resource Description Framework) or ontology languages like OIL (Ontology Inference Layer).

Integrity constraints definition in an ontology

For mobile data acquisition not only restrictions and inconsistent relations have to be considered, but also obligatory relations within the data model and quality information related to the data. Adequate tests have to be done before newly acquired data is transferred to the server. But since the necessary information is not contained in the available XML-schema, some other kind of formalization method is needed. In our approach, we refer to the ontology of the data model. An ontology contains the totality of geospatial concepts, categories, relations and processes of the data model as well as their interrelations. It captures the semantic of information contained in the data model. Compared to a schema, ontologies are representing concepts in the real world, while a schema only refers to those parts of the concepts, which are stored in a database. These differences between an ontology and a schema are extensively described in (Fonseca et al. 2003). They pointed out that a conceptual schema leaves out some of the concepts and ideas the data modeler and the later user have agreed upon. The same applies for the physical schema we are working with. For quality assurance, this background knowledge of the user is very important, since it implies rules and restrictions on the object classes, which have to be added to the schema information. As shown in (Mostafavi et al. 2004) an ontology in principle can be used for quality assessment of spatial databases. But up to now there is no formalism known to describe spatial data integrity constraints.

Definition of integrity constraints in OWL

In our approach, we investigated the possibilities of OWL (Web Ontology Language) for the description of spatial data integrity constraints. OWL is an XML (eXtensible Markup Language) and RDF (Resource Description Framework) based semantic markup language for publishing and sharing ontologies (W3C 2004). It has to be pointed out that our aim is not to define an exhaustive application ontology. Our focus is to investigate the possibilities for defining constraints and integrity rules within an ontology and using this as an extension to the existing data schema. Figure 4 schematizes the information separation between the XML-schema and the OWL-ontology.

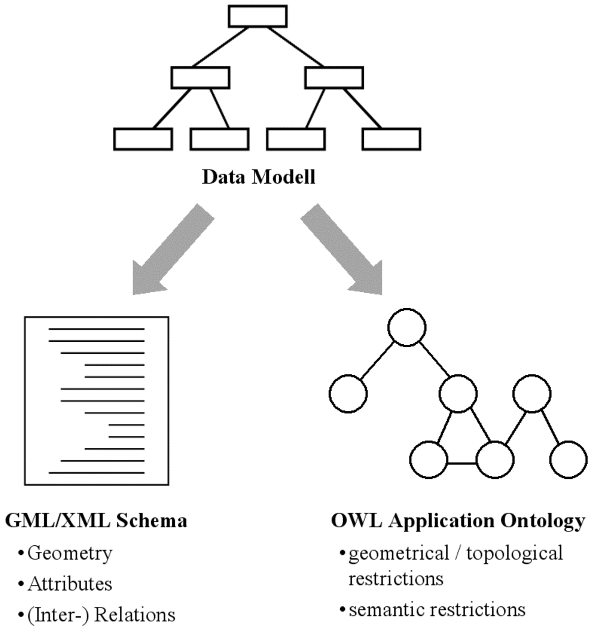


Figure 4: Separation of information between schema and ontology

To enable the definition of integrity constraints in OWL, the relations (see *Figure 3*) have to be treated as object properties of the relevant object classes. For the semantic description of such properties, OWL provides axioms like subPropertyOf, symmetric, transitive or inverseOf. Axioms are prescriptions for the complement of semantic knowledge. They represent implicit knowledge about concepts i.e. objects and relations. Through the complemented axioms it is possible, to design a hierarchy of spatial relations (*see Figure 5*), which later on can also allow for reasoning. In this hierarchy, for example, the property "Equal" is defined as a symmetric, transitive, sub property of "Intersects" and the inverse property of itself. "subPropertyOf " means in this case: if two objects are equal they are also intersected. The generic "Spatial Relation" on top of the hierarchy is abstract. The denotation and the definition of the spatial relations refer to the spatial operators defined in OGCs Filter Encoding Implementation Specification (OGC 2001). Further explanations on the employed axioms and examples for corresponding XML formalizations are given in (W3C 2004). An Example for logical reasoning based on that hierarchy is the conclusion from the fact "a clearing is within a certain forest" to the inverse "the forest is containing the clearing". This is possible because "Within" and "Contains" are defined as inverse properties of each other.

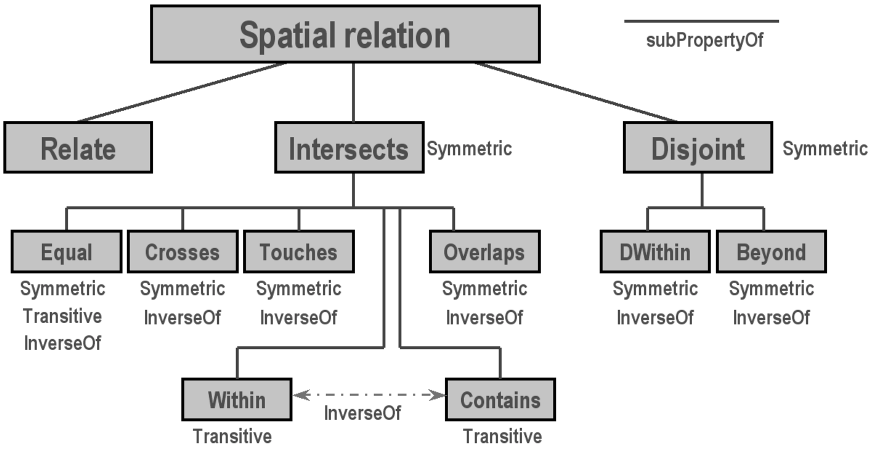
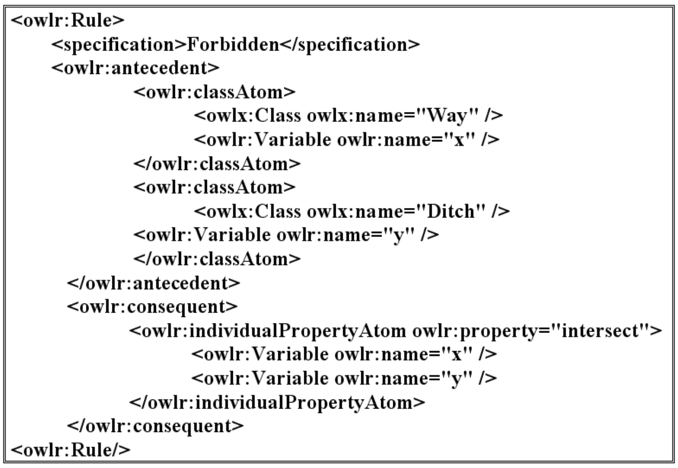


Figure 5: Hierarchy of spatial relations

Unfortunately, OWL doesn't provide the possibilities for a more extensive definition of object properties through axioms. This fact has been exposed in more detail by (Horrocks et al. 2004). They pointed out the limitations of OWL in terms of the axioms and proposed an OWL Rules Language (OWLR) as a syntactic and semantic extension to OWL. OWLR introduces rules as a new kind of axiom. These rules consist out of an antecedent and a consequent, each one containing zero or more atoms. Atoms contain conditions like data literals or variable assignments (among others). Multiple atoms in the antecedent are treated as a conjunction. Therewith it is possible to define rules on object properties as well as arithmetic relationships between data values of certain attributes. In our approach, we decided to use OWLR because it provides more expressive possibilities to encode constraints. In ontologies, integrity constraints should be treated as axioms, which is not supported in a convenient way by OWL. Complex rules referring to multiple object classes, their attributes and interrelations are not possible at all. Since OWLR is tightly integrated into OWL, its extensions might be considered in future development of OWL.

Figure 6 shows an example of a topological constraint (comparable to those defined by (Ubdea & Egenhofer, 1997)) encoded in OWLR. The variables x and y are assigned to the object classes way and ditch in the antecedent and used as argument for the “intersect” property in the consequent. The specification determined in the first sub element refers to the definition made by Ubdea and Egenhofer. The informal meaning of it is: an intersection of a way and a ditch is forbidden. Generally the rules can be read as: if the conditions in the antecedent hold then the conditions in the consequent must also be meet.



```
<owlr:Rule>
      <specification>Forbidden</specification>
      <owlr:antecedent>
                  <owlr:classAtom>
                           <owlx:Class owlx:name="Way" />
                           <owlr:Variable owlr:name="x" />
                  </owlr:classAtom>
                  <owlr:classAtom>
                           <owlx:Class owlx:name="Ditch" />
                  <owlr:Variable owlr:name="y" />
                  </owlr:classAtom>
        </owlr:antecedent>
        <owlr:consequent>
                 <owlr:individualPropertyAtom owlr:property="intersect">
                          <owlr:Variable owlr:name="x" />
                          <owlr:Variable owlr:name="y" />
                 </owlr:individualPropertyAtom>
        </owlr:consequent>
<owlr:Rule/>
```

Figure 6: Example for a rules definition in OWLR

One of the main advantages of this structure is, that it can also be applied for arithmetic and other relations (instead of the topological relation) and attributes, respectively their values (instead of the object classes). Even complex constraints, which are restricting various object types, their spatial relation and attribute values are possible.

Such semantic description and rules can be easily applied directly within the data acquisition process. Possible conflicts between the entered attribute values or the acquired geometry and the geometry of already existing objects are recognized during the acquisition in the field. So it is assured that the producer’s perception of the data model is correctly attended to, even by inexperienced users. Newly recorded data can be evaluated during the acquisition process using semantic knowledge about the data model. This assures that the collected data is conform to the underlying data model and achieves all quality requirements before it is transferred to the server.

APPLICATION SCENARIO

The presented architecture can be flexibly applied in different application domains. One contingent application is in the field of landslide monitoring (Kandawasvika et al. 2004). Proposed test areas are in Balingen and Rosenheim, Germany. These areas have unstable surfaces in which rock masses, soil and other loss material may fall from the mountain slopes and be hazardous to people using the nearby roads or walking paths and to other infrastructures close by. Since these incidents can’t be predicted, measurements of any surface movements are done by on-site extensometers, installed in cracks, gaps or ditches in the active landslide parts of the mountains. If the observations extend certain values an alarm is automatically triggered. For manual measurements, especially to validate these alarms, also other instruments, like total stations, GPS receivers, normal tapes, laser distance meters etc., are used by the fieldworker.

The proposed system supports the whole workflow of the geologists during the field visit. Existing information about geological objects and measurement devices can be downloaded and visualized together with the users position to assist the orientation in the environment. Previous measurements can be analysed and compared to the manual observations for verification. If any new ditches or other geological phenomena appear, the user can select the corresponding object class from the data schema and thereupon acquire the geometry and related attributes based on that schema. Through a network of WLAN access points the mobile client stays connected to the databases and can directly insert or update data.

For the landslide application scenario, a data model containing all types of geological phenomena occurring in the test area and their relations has been developed. This model is particularly adjusted to the situation of geologists validating an alarm in the field and acquiring newly emerged geological objects. Thereby the defined integrity constraints can not only help to assure the consistency of the database. Using the integrity constraints the application can also advert to dangerous situations that are identified in the data but eventually not directly visible in the field. For example, if the fieldworker is measuring a newly arisen ditch, which is within a close distance to a publicly accessible way, the system can automatically detect a narrow footpath and tell the fieldworker to close it for the public use.

CONCLUSIONS

In this paper we give an overview on the application of mobile online data acquisition and the corresponding quality assurance process. In detail we give a proposal for a framework for the formalization of integrity constraints in OWLR (Horrocks et al. 2004), an extension of the semantic markup language OWL (W3C 2004). The paper is showing how integrity constraints can be defined in an ontology and how these constraints can be used during mobile data acquisition as an extension to the data model described by a schema. The constraints define conflicts between attribute values or the acquired geometry and the geometry of already existing objects. Therewith a violation against rules like obligatory topological relations and inconsistent interrelations with the already existing data is recognizable during the acquisition in the field.

In future, the potential of RuleML, an XML-based markup language for rules currently under development, as well as possible extended versions of OWL have to be considered for the constraint definition. For our approach, it will be investigated how the proposed constraint specifications can be extended by values like for example a severity, which also allows for the treatment of constraints, that are not so strict. The application of the constraints and their integration into the workflows will be checked in on-going field tests.

ACKNOWLEDGEMENTS

The funding of the research project “Advancement of Geoservices“ (Weiterentwicklung von Geodiensten) by the German Ministry of Education and Research (BMBF) under grant no. 03F0373B et al. within the framework of the GEOTECHNOLOGIEN initiative (http://www.geotechnologien.de) is greatfully acknowledged. The responsibility for the contents of this publication is by the authors. This is publication no. GEOTECH-116 of the programme GEOTECHNOLOGIEN.

BIBLIOGRAPHY

Breunig, M.; Malaka, R.; Reinhardt, W.; Wiesel, J.(2003): Advancement of Geoservices. GEOTECHNOLOGIEN Science Report No. 2, Information Systems in Earth Management, Potsdam, pp. 37-50.

Breunig, M.; Häussler, J.; Reinhardt, W.; Wiesel, J.(2004): Developing Mobile Spatial Services for the Geosciences. In: 19th International CODATA Conference, Berlin, Proceedings, 2004

Cockcroft, S. (2004): The Design and Implementation of a Repository for the Management of Spatial Data Integrity Constraints; In: GeoInformatica 8:1 2004, pp. 49-69, Kluwer Academic Publishers 2004

Fonseca, F.; Davis, C.; Câmara, G.(2003): Bridging Ontologies and Conceptual Schemas in Geographic Information Integration, In: GeoInformatica 7:4 2003, pp. 355 - 378, Kluwer Academic Publishers 2003

Horrocks, I.; Patel-Schneider, P.(2004): A Proposal for an OWL Rules Language. In: Proceedings of the 13th international conference on World Wide Web, pp 723 - 731, 2004, ACM Press NY

Kandawasvika, A.; Plan, O.; Reinhardt, W.; Mäs, S.; Wang, F.(2004): Concepts and Development of a Mobile Client for Geospatial Data Acquisition. In: GeoLeipzig 2004 - Geowissenschaften sichern Zukunft. Schriftenreihe der Deutschen Geologischen Gesellschaft Heft 34. 2004, p.79.

Kandawasvika, A. & Reinhardt, W. (2005): Concept for interoperable usage of multi-sensors within a landslide monitoring application scenario. Submitted paper for 8th AGILE Conference on GIScience, 2005, Estoril, Portugal; unpublished, 2005

OGC (2001): Filter Encoding Implementation Specification, Version: 1.0.0, OpenGIS® Implementation Specification, OpenGIS project document: OGC 02-059, 19 September 2001

OGC (2002a): Web Feature Service Implementation Specification, Version 1.0.0, OpenGIS® Project Document: 02-058; 19 September 2002

OGC (2002b): Geography Markup Language (GML) Implementation Specification, Version 2.1.2, OpenGIS Project Document Number 02-069, 17 September 2002

Mostafavi, A.-M.; Edwards, G.; Jeansoulin, R. (2004): An ontology based Method for Quality assessment of spatial data bases. In: ISSDQ '04 Proceedings; Frank, A. and Grum, E. Vol.1,28 a,b; 2004

Plan, O.; Mäs, S.; Reinhardt, W.; Kandawasvika, A.; Wang, F.(2004): Konzepte für die mobile Erfassung von Geodaten. In: Geoinformation und Mobilität von der Forschung zur praktischen Anwendung, Tagungsband Münsteraner GI-Tage 2004, IFGI prints. 2004, p.97-105

Pundt, H. (2002): Field Data Collection with Mobile GIS: Dependencies Between Semantics and Data Quality, In: GeoInformatica 6:4 2002, pp. 363-380, Kluwer Academic Publishers

Sayda, F.; Reinhardt, W.; Wittmann, E.(2002): GI and Location Based Services for Mountaineers. In: Geoinformatics, Magazine for Geo-IT Professionals, 2002, Nr. 5.

Servigne, S.; Ubeda, T.; Puricelli, A.; Laurini, R. (2000): A Methodology for Spatial Consistency Improvement of Geographic Databases, In: GeoInformatica 4:1 2000, pp.7-34, Kluwer Academic Publishers 2000

Ubeda, T.; Egenhofer, M. J.(1997): Topological Error Correcting in GIS. In: Advances in Spatial Databases, Fifth International Symposium on Large Spatial Databases, SSD'97, Berlin; Lecture Notes in Computer Sience, Vol 1262, Springer Verlag, pp. 283-297, 1997

W3C (2004): OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004
Editors: Dean, M. & Schreiber, G.
Available at: http://www.w3.org/TR/2004/REC-owl-ref-20040210/