Ontology Based Quality Assurance for Mobile Data Acquisition

Fei Wang, Stephan Mäs, Wolfgang Reinhardt and Admire Kandawasvika [1]

Abstract

Mobile data acquisition provides a new chance for the fieldwork in geosciences. In this respect, advances in ICT are enabling effective and efficient access to and acquisition of geodata in field. Within the project "Advancement of Geoservices" a mobile data acquisition system has been built up based on OGC's Transactional Web Feature Service (WFS-T) and extensions. Therewith using this system mobile field workers can not only have interoperable access to heterogeneous databases in the field, but also use transaction operations like inserting, deleting and updating data. The acquisition system is generic and can be easily adapted to different geosciences applications. The direct transfer of newly collected data into databases requires data quality assurance directly in the field. In the project mentioned above, the overall data quality assurance procedure has been investigated. In this paper, we focus on one part of this procedure, which addresses the question how ontology-based rules can be incorporated to consider user defined quality constraints. Semantic Web Rule Language (SWRL), a member submission of W3C is used to define the data quality constraints. Mobile geocomputing and quality checking tools have been created and integrated into the mobile data acquisition workflow. These tools can simultaneously check the data measured or entered by field workers against the defined ontology-based data quality constraints. Therefore the inconsistencies between collected data and data model can be avoided at an early stage. Mobile field workers can benefit from the proposed way in many aspects, which are presented in this paper.

Keywords: data quality, ontology, mobile data acquisition, OWL, SWRL

1. Introduction

Environmental monitoring and risk management requires large amount of accurate and up to date data, which are often collected during field surveys. The rapid development of the mobile technologies now allows for new approaches during the fieldwork as follows: advanced portable computers (e.g. rugged tablet PCs and other handheld devices) are available for field data collection; OGC's WFS-T provides a way to retrieve the geospatial datasets from heterogeneous databases, and it also allows for the transaction operations like inserting, deleting and updating data backwards to the databases; Mobile communication technologies like WLAN, GPRS and UMTS give people opportunities to transfer the data from the field back to the database. Especially UMTS and WLAN have the possibilities to support the fast data rates which are quite useful for mobile users in order to save time when transferring larger amounts of data in the field, and their interworking can achieve better performance (Schmidt, 2004). Therewith, field workers don't need to synchronise all of the necessary data in the office before going to the field.

Mobile data acquisition enables the inclusion of functions that support the improvement of data quality during outdoor data collection. One approach to support field workers during data collection is the ontology-based capture of data, which is based on a list of rules that can provide diagnosis for the user to make decisions. These plausibility checks enable the user to correct errors in the field, thus guaranteeing a higher quality of data (Pundt, 2002). As parts of the data model, spatial, topological and semantic quality constraints and combination of them are defined to express the plausibility rules. Besides those constraints,

[1] AGIS – GIS lab, University of the Bundeswehr Munich Werner-Heisenberg-Weg 39 D-85577 Neubiberg, Germany
Email: (fei.wang,stephan.maes,wolfgang.reinhardt,admire.kandawasvika)@unibw-muenchen.de, Internet: http://www.agis.unibw-muenchen.de

the instruction information about how users should react to the possible errors is also defined. In this paper, the definition of the quality constraints is introduced and a second emphasis is put on the integration of them into the mobile acquisition workflow.

The paper is organized as follows: Section 2 introduces the application scenario. Section 3 and 4 explain how the data quality constraints and instructions can be defined in the ontology language SWRL and how they are integrated into the data acquisition workflow. It also shows how the mobile field users can benefit from the proposed quality assurance method. Section 5 describes the results of the first field tests. Finally section 6 gives a conclusion of this paper.

2. Application scenario in landslide areas

Figure 1
On-site extensometers in the test area of Balingen, Germany

Within the project “Advancement of Geoservices” a prototype for a mobile data acquisition system has been developed which is particularly designed for an application in landslides monitoring and decision making. In this specific scenario the application supports monitoring of geological phenomena and acquisition of geological objects like ditches and edges. In cooperation with local authorities, two test areas have been selected. One of the test areas is located near Balingen, Germany. This area has environmental problems of landslides. Due to slope instability, if unstable rock masses, soil and other loss material fall from the mountain slopes, it might become very dangerous for the citizens and nearby infrastructure. In order to get immediate warning information about the slope movements, permanent survey instruments have been installed by local authorities, for example, the on-site extensometers shown in Figure 1 (two separate extensometers are installed at different places), which are normally installed in cracks, gaps or ditches in the sensible areas, and automatic total stations. In case of significant surface movement these instruments can automatically trigger an alarm, which is then sent to the control station.

In this case the geologists have to go to the field in order to validate the alarm information in the field, because measurement instruments like extensometers are also very sensible to the disturbances from animals or falling objects. The developed mobile data acquisition system enables the field workers to fulfil the on-site tasks easily. Some examples of its functionality are: the mobile acquisition system can be connected to remote databases through GPRS, UMTS or WLAN network [refer to chapter 5], a map of the current situation as well as previous measurements can be downloaded and displayed for example on a tablet PC, and the geometry and the attributes of new features can be acquired by different types of geodetic sensors like GPS, total stations, laser scanners, extensometers, etc. (Kandawasvika and Reinhardt, 2005). Quality validation is integrated into those steps according to defined quality constraints. Detailed system introduction and its functionality can be referred to (Plan et al., 2004).

3. Ontology based quality constraints definition

Data quality is always one of the key issues in any GIS. During mobile data acquisition, it plays an important role since all of the collected data in the field are supposed to be immediately transferred to the databases. In order to ensure data quality and reduce error risks, a quality assurance method integrated into the mobile acquisition workflow has to be taken into account.

In this paper we put more emphasis on a method that provides the mobile data acquisition system with constraints as an extension to the data application schema. In that conjunction we have to mention that OGC's WFS-T is used to provide the mobile system with the application schema, moreover the transactional services of WFS-T can be used to transfer newly collected features to a remote server. For such task it is well known that the WFS-T server is able to provide an XML schema document according to the feature types listed in the request. This XML schema document is a GML application schema that can be used to validate collected feature instances which should be sent to the remote server. However the application schema only includes information about the features, e.g. geometry type, attribute name, attribute data value type of and etc. With this information the data can only be assured with regard to geometry type, attribute fields and data types. But from the users point of view there are more requirements especially with regard to data integrity which should also be checked in the field. For example, a topological constraint of the mentioned landslide scenario "a hiking way is not allowed to be intersected with a ditch" can not be provided by the normal GML application schemas. Therefore a way of extending the GML application schema has to be investigated.

In our approach an ontology based method is used for that. This idea has been introduced earlier in (Mostafavi et al., 2004) and (Pundt, 2002). The selection of a certain ontology formalization language for the definition of quality constraints depends primarily on its expressive power. SWRL member submission document of W3C (W3C, 2004), which bases on the combination of sublanguages of Web Ontology Language (OWL) and Rule Markup Language (RuleML) gives a new chance for the definition of logical relationships in an ontology language. Therefore in this research work, the Semantic Web Rule Language (SWRL) is attempted for defining data integrity constraints. Topological and semantic constraints as well as their combinations are defined in the ontology language to ensure the data quality. For the definition of such constraints, spatial relations like "intersect" or "disjoint" have to be used. Therefore a more detailed description is given in (Mäs et al., 2005a). In SWRL a rule axiom consists of an antecedent (ruleml:body) and an consequent (ruleml:head). Informally a rule may be understood by the meaning that if the antecedent holds (is "true"), then the consequent must also hold. An example to how the concept is implemented in SWRL is given in Figure 2.

The annotations in line 2 to line 5 of the example enable for a further description of the constraint. The "constraintID" item contains the index of this constraint. In our definition "severity" value can have three different values: "strict", "avoid violation" and "apply with caution, users reaction necessary". The first one means that a violation of the constraint is illegal and the violating data has to be changed. The two other values leave it up to the user's decision with respect to what has to be done in case of a violation. Therewith it is possible to use constraints as a description of (maybe unusual) relations of objects, which are not strictly forbidden but nevertheless have to be checked. The third value additionally requires some reaction by the user, e.g. the user should record the current situation according to the real world environment for the other possible users. The “comment” and the “correctionInstruction” items provide users with helpful information about how to react to the violation.

```
1  <ruleml:imp>
2      <swrlagis:constraintID rdf:datatype="http://www.w3.org/2001/XMLSchema#positiveInteger">2</swrlagis:constraintID>
3      <swrlagis:severity>to apply with caution, users reaction necessary</swrlagis:severity>
4      <rdfs:comment>Ways that are publicly accessible should not be closer to ditches than 10 Meters</rdfs:comment>
5      <swrlagis:correctionInstruction>Set way attribute "publicly accessible" to false, close the way for public access </
   swrlagis:correctionInstruction>
6      <ruleml:_body>
7          <swrlx:classAtom>
8              <owlx:Class owlx:name="Way"/>
9              <ruleml:var>way</ruleml:var>
10         </swrlx:classAtom>
11         <swrlx:classAtom>
12             <owlx:Class owlx:name="Ditch"/>
13             <ruleml:var>ditch</ruleml:var>
14         </swrlx:classAtom>
15         <swrlx:individualPropertyAtom  swrlx:property="publiclyAccessible">
16             <ruleml:var>way</ruleml:var>
17             <ruleml:var>x</ruleml:var>
18         </swrlx:individualPropertyAtom>
19         <swrlx:builtinAtom  swrlx:builtin="swrlb:equal">
20           <ruleml:var>x</ruleml:var>
21           <owlx:DataValue owlx:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</owlx:DataValue>
22         </swrlx:builtinAtom>
23     </ruleml:_body>
24     <ruleml:_head>
25         <swrlagis:builtinAtom swrlagis:builtin="dWithin">
26             <ruleml:var>way</ruleml:var>
27             <ruleml:var>ditch</ruleml:var>
28             <swrlagis:specification>Forbidden</swrlagis:specification>
29             <swrlagis:distance>10m</swrlagis:distance>
30         </swrlagis:builtinAtom>
31     </ruleml:_head>
32 </ruleml:imp>
```

Figure 2
Example of quality constraints defined in SWRL

Line 6 to line 23 contain the antecedent part that shows the assignment of two variables *way* and *ditch* as Way and Ditch objects, and the Boolean data type attribute “publiclyAccessible” (which means whether the way is closed for public access or not depending on its availability for walking on) of the *way* is true. Line 24 to line 31 present the consequent part that defines a relation between these two spatial objects. The “dWithin” item is a spatial relation which means two spatial objects disjoint with each other within a certain value. The “specification” item used in the head part refers to the definition by (Ubeda and Egenhofer, 1997), and can have the possible values “Forbidden” (i.e. zero times allowed), “At least n times”, “At most n times” or “Exactly n times”. The explanation of this example in possible landslide scenario and how to integrate this example into the data acquisition workflow will be explained in the next chapter.

A more detailed introduction about how to construct SWRL rules can be found in (Mäs et al., 2005b). The definition of rules is not restricted to relate only two object classes or attributes. Even complex spatial and topological relations between numerous spatial objects together with their attribute values can be described. Because of the ontology based quality constraints are also based on XML structure, they can be easily attached to GML application schema. The constraints are encoded as annotations to each GML application schema with respect to their corresponding feature classes. Therewith the quality information defined in the constraints is transferable and available for the users during the data acquisition workflow, and the quality assurance task can be implemented based on that.

4. Quality assurance integrated into the data acquisition workflow

As mentioned in the last section, the description and the constraints of the data model are defined in SWRL and used as an extension to GML application schemas. Therewith newly collected data in the field can be checked against the defined constraints, and the errors can be recognized during the data acquisition workflow.

For the quality assurance an independent Java based quality assurance package (QA) has been developed and integrated into the mobile data acquisition software as a component. The introduction of the architecture of the acquisition software refers to (Häußler et al., 2004). The QA gets the GML application schema and automatically maps the schema elements into java objects. Such mapping technology was introduced in (Kalyanpur et al., 2004). The java objects include the information about the constraints which are used to validate the data captured by the mobile field users. To handle geometrical and topological operations the QA uses JTS Topology Suite, which is an Open Source Java API that implements functions of spatial data model and spatial predicates of the OpenGIS Simple Features Specification For SQL Revision 1.1 (OGC, 1999). It provides many common operations for computational geometry and spatial data processing. For a detailed introduction about how to use JTS Topology Suite refers to (Aquino, 2003).

According to the constraint given in Figure 2, the way how QA serves the mobile data acquisition software for automatically checking the measurement data is given in a possible scenario:

A geologist comes to the landslide area and wants to capture the new ditches in the field. Firstly he has to set up the local network and install the proper measurement instrument on the site, like GPS receiver or total station for data collection. Then he gets the mobile acquisition software GUI with map of surrounded areas in the tablet PC as shown in Figure 3, and configures the measurement instrument plugin (component to control the instrument) from the GUI. Through the network connection, the QA loads the GML application schema from WFS server and maps it into java objects. Therefore the QA can get the constraint information of Figure 2 for quality assurance task in next step.

When the geologist has measured the geometry of a new *ditch*, the QA automatically detects that a *way* is within a distance of 10 meters of the *ditch* and the "publiclyAccessible" attribute of this *way* is true, and pops up a warning window with information "*Ways that are publicly accessible should not be closer to ditches less than 10 meters. Set the way attribute publiclyAccessible to false, Way should be closed for public access.*" to the user as given in Figure 3. After the violation has been indicated it is up to the responsibility of geologist how to react, because the minor severity of the constraint doesn't force a change of the data. In case there is a fence in between the ditch and the way he might rate the situation as harmless. So the geologist only has to document this decision for future checks. In case there is no fence or the fence is not firm enough, he might decide to set the "publiclyAccessible" attribute to false and put the "danger" signs to block the way.

The previous example constrains the spatial relation (dWithin) of *way* and *ditch* objects as well as semantic aspects of the *way* attribute "publiclyAccessible". Since a validation against this constraint requires the attributes and the geometry of the features, the checking is done at the end of the acquisition, when the whole feature is acquired. Other constraints might only consider spatial relations between features or a set of attributes of a single feature. Since the acquisition of geometry and attributes is handled separately in our application, the checks against such constraints can be done directly after that respective part of the acquisition is finished. Therewith mistakes are recognized as early as possible during the acquisition workflow.

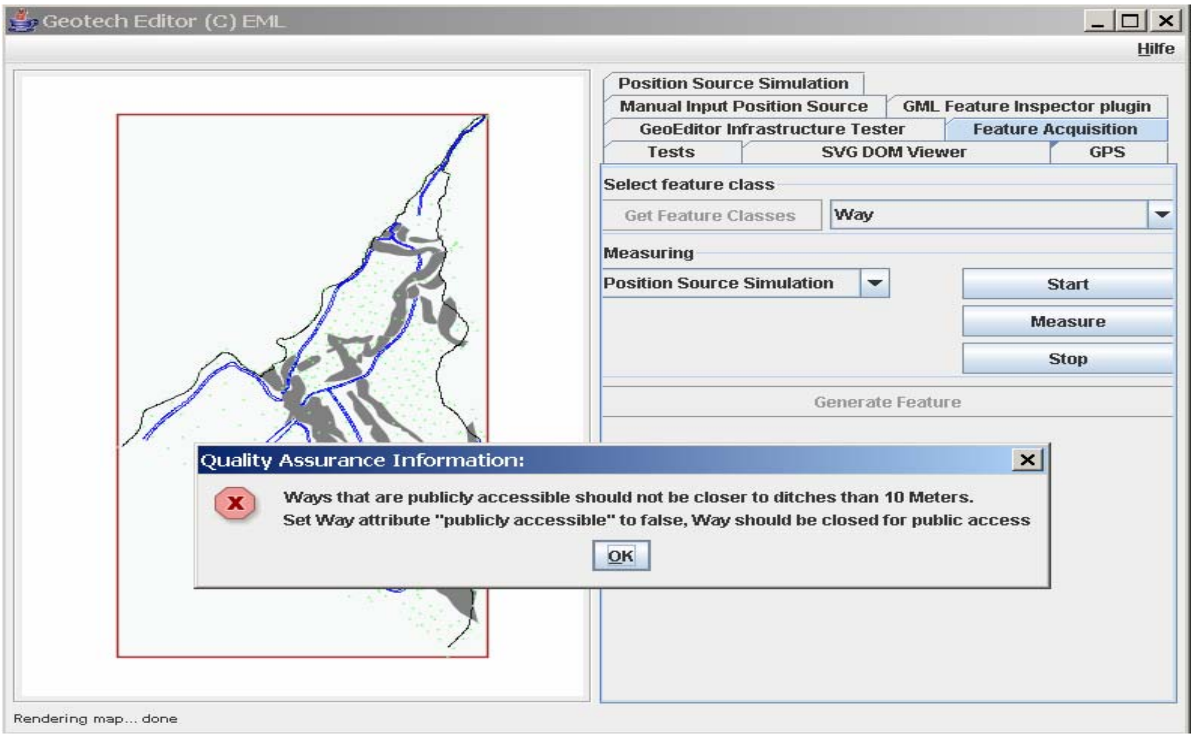

Figure 3
GUI of mobile data acquisition

When all of the defined constraints are checked during user's data collection steps, the new feature can be displayed on the map like on the left panel of GUI in Figure 3.

The example above describes how data quality can be checked in the field through the ontology-based integrity constraints and the rule-based functions that are integrated into the mobile data acquisition system. Therefore the data quality of collected data can be guaranteed through the plausibility checks, and users are also confident about their actions.

5. Field Tests

In order to evaluate the mobile data acquisition system, field tests were carried out in May 2005 in landslides area of Balingen, Germany. According to the natural environment of the test area which has high density of tall trees and very steep terrain, the primary aims for this field tests involve investigations, e.g. WLAN accessibility and performance of mobile acquisition workflow including quality assurance task.

The field tests system configuration is shown in Figure 4. For the first field tests a field server with geodatabase was installed, and a local WLAN network which could be connected to a remote server was set up, so that client application running on a tablet PC could access the field server through the WLAN connection. Because of the high density of tall trees with plenty of branches and leaves in May, GPS availability was not good. This test has to be repeated with new generation of GPS receivers. The forest covers an area of around 150*250meters. To cover the whole area with WLAN only two bridge access points had to be installed. We assume that this very good signal range of up to 150 meter was mainly caused by signal reflections of crown canopy in the forest.

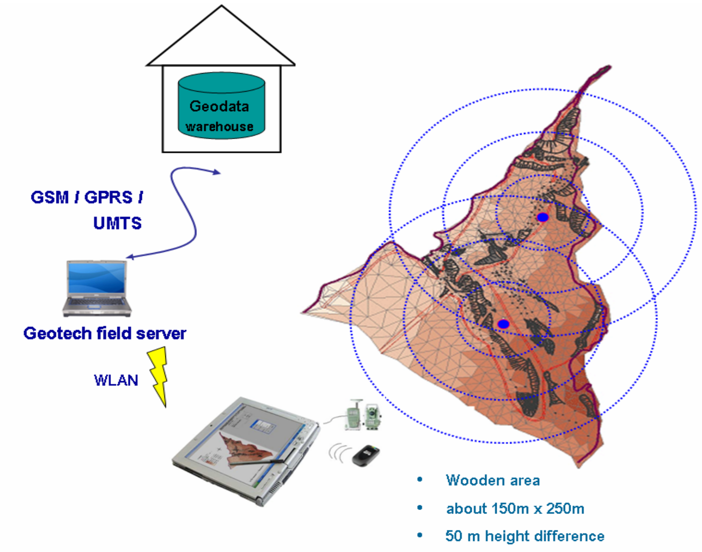

Figure 4
Balingen field test system configuration

The whole mobile data acquisition was tested with a total station. New measurement was collected and the quality assurance package worked as explained in Chapter 4. Field user could receive immediate error or warning information according to the predefined quality constraints.

These first field tests have shown that:

- the mobile users could get the surrounding environment data;
- the mobile users had stable connection with the field server;
- the quality assurance could be done through the plausibility checks and the correction instruction;

In the near future, more measurements instruments and mobile communication like GSM, GPRS or UMTS will be tested, professional GPS receiver will be adopted, and different geosciences application fields will be taken into consideration.

6. Conclusions

In order to ensure the data quality of collected data based on the mobile data acquisition, an ontology based quality assurance method is proposed in this paper. Ontology based data modelling and definition of quality constraints using SWRL is applied. The description of quality constraints is integrated with WFS-T GML application schemas so that the quality information is accessible and sharable via the Internet. The developed Java-based quality assurance package used the defined quality constraints. Therefore quality plausibility checks are integrated into mobile acquisition workflow. Mobile field users can get immediate error or warning information together with instruction, which are displayed on the mobile acquisition graphical user interface in a tablet PC, when exceptions happen. The first field tests have proven the feasibility and usefulness of the integrated data quality assurance method.

The approach in the research can be adopted into many GIS data quality assurance applications. SWRL is still under development, so updates of this language should be taken into account in future.

7. Acknowledgement

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-130 of the programme GEOTECHNOLOGIEN.

Bibliography

Aquino, J., (2003), JTS Topology Suite Developer's Guide, VIVID SOLUTIONS.

Guptill, S. C., and Morrison, J. L.eds., (1995), Elements of Spatial Data Quality, Elsevier.

Häußler, J., MERDES, M., and ZIPF, A., (2004), A Graphical Editor for Geodata in Mobile Environments, in G. S. Report, ed., Information Systems in Earth Management, Potsdam, p. 55-58.

Kalyanpur, A., Pastor, D., Battle, S., and Padget, J., (2004), Automatic Mapping of OWL Ontologies into Java: Proceedings of Sixteenth International Conference on Software Engineering and Knowledge Engineering (SEKE).

Kandawasvika, A., and Reinhardt, W., (2005), Concept for interoperable usage of multi-sensors within a landslide monitoring application scenario: Proceeding 8th AGILE Conference on GIScience.

Mäs, S., Reinhardt, W., Kandawasvika, A., and Wang, F., (2005a), Concepts for quality assurance during mobile online data acquisition: Proceeding 8th AGILE Conference on GIScience.

Mäs, S., Wang, F., and Reinhardt, W., (2005b), Using Ontologies for Integrity Constraint Definition: 4th International Symposium on Spatial Data Quality, unpublished.

Mostafavi, M. A., Edwards, G., and Jeansoulin, R., (2004), An ontology-based method for quality assesement of spatial data bases: ISSDQ'04 Proceedings.

OGC, (1999), OpenGIS Simple Features Specification For SQL, Revision 1.1, OpenGIS Project Document 99-049.

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

Pundt, H., (2002), Field Data Acquisition with Mobile GIS: Dependencies Between Data Quality and Semantics: GeoInformatica, v. 6, p. 363-380.

Schmidt, A. L., (2004), UMTS and WLAN interoperability: Master's thesis, Technical University of Denmark.

Ubeda, T., and Egenhofer, M. J., (1997), Topological Error Correcting in GIS: Springer Verlag , Advances in Spatial Databases, v. 1262, p. 283-297.

W3C, (2004), SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission.