USING SDI AND WFS FOR QUALITY ASSURANCE ON FIELD DATA COLLECTION

Mohammad Hosseinpour[1], Ali Asghar Alesheikh[2], Majid Hamrah[2]

[1]GIS M.Sc student at KNT University of Technology   Hosseinpour. - mhp@gmail.com
[2]Associate Professor at KNT University of Technology
- Alesheikh@kntu.ac.ir

KEY WORDS: SDI, Web Feature Services, Data Quality, Field Data Collection

ABSTRACT:

Nowadays, various organizations collect new spatial data or update existing ones. If a specific information community collects spatial data with regards to its own requirements, purposes and applications then other information communities may not be able to use the data easily. In addition to quality issues of the collected data, the semantic heterogeneity plays a major role in spatial data sharing. An important objective of establishing spatial data infrastructure (SDI) is to share data between diverse information communities through exerting policies and standards. As such, SDI can provide a suitable framework to for collecting, sharing and using data.Web Feature Services (WFS) are suitable tools for achieving this aim because they consist of information about features definitions and their complete attributes. This leads to solve the problem of different understandings from a specific feature. In addition to this, it is possible to formulate integrity constraints using ontology programming languages and implement the rules during field data collection.This paper aims to investigate the way of imposing quality controls on acquired data during field data collection. As a prototype, a mobile data acquisition system has been designed that warns users when any data inconsistency occurs and users can remove the error from database with correcting the occurred error instantly.

1. INTRODUCTION

Nowadays, various organizations collect new spatial data or update existing ones. If a specific information community collects spatial data with regards to its own requirements, purposes and applications then other information communities may not be able to use the data easily. In addition to quality issues of the collected data, the semantic heterogeneity plays a major role in spatial data sharing. An important objective of establishing spatial data infrastructure (SDI) is to share data between diverse information communities through exerting policies and standards. As such, SDI can provide a suitable framework for collecting, sharing and using data.

At the other side, in the field of data acquisition, using mobile geographic information system (Mobile GIS) has facilitated the task of collecting or updating data and whereas these data are mostly transferred directly to database so there is need to assess them from the quality viewpoint and semantic accuracy to avoid inconsistency in database. The best way for this work is to control data during their collection in the field.

The aim of this paper is to address the way of exerting these quality controls on data during their gathering in the field. In order to achieve the mentioned scope, the basic concepts like SDI, WFS and Mobile GIS are described at first and then in the following sections these concepts and technologies are used to impose quality controls from the early stages of data acquiring and in the last section an implemented mobile data acquisition system is presented as a prototype.

2. SPATIAL DATA INFRASTRUCTURE

Spatial data refers to the subset of data that is related to the earth such as: topographic data, geographic features and height information**.** Approximately 80% of used data in government decisions are spatial or at least related to the earth [1]. For using spatial data, they must be accessible and easily usable. These data have a major role in spatial decisions and people can have a better decision with using suitable spatial data.

Many private and public organizations are acting in the field of producing and maintaining spatial data and their products perhaps are the same. Then for avoiding these extra works and saving money and time, data sharing will be the crucial issue in the field of spatial data. That means each organization that produces a dataset, should inform users and other organizations that may need that data.

1.1 SDI Definition

There are some definitions for spatial data infrastructures but the thing that is common between them is creating the area that all could collaborate with each other to reach their purpose in all organizational and political levels. Regarding the different existing branches, there are different definitions for SDI but in this paper Groot and McLaughlin’s definition of Spatial Data Infrastructure has been adopted [2]:

“Spatial Data Infrastructure encompasses the networked spatial databases and data handling facilities, the complex of institutional, organizational, technological, human, and economic resources which interact with one another and underpin the design, implementation, and maintenance of mechanisms facilitating the sharing, access to, and responsible use of spatial data at an affordable cost for a specific application domain or enterprise.”

2.2 Core components of SDI

With reference to existing definitions for SDI, the main components of SDI can be recognized. As shown in Figure 1 these components are: people, access network, policy, standards and data.

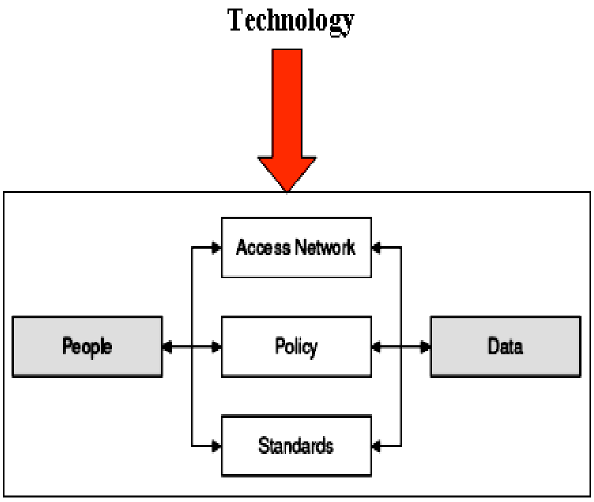

Figure1. SDI components (Rajabifard & Williamson 2001 p.5)

Each of the five components is described briefly below [4]:

The *data* component comprises the core data elements for the SDI. For example a state SDI could define geodetic control, cadastral information, administrative boundaries, elevation and hydrology data themes as fundamental.

The avenue by which data within an SDI is made available to the community can be described as the *access network*. Access arrangements must be made in accordance with the policy decisions and technical specifications defined within the implementing organization's institutional framework.

The institutional framework of the organization implementing the SDI defines the *policy* and administrative arrangements for collecting, maintaining, accessing and applying the standards and data sets.

The *standards* component defines the technical characteristics of the fundamental data sets. These can include metadata, data dictionaries, data quality, data transfer, reference systems and data models.

The *people* component of an SDI encompasses the diversity of the users and producers of spatial data (including value-adding agents).

These components can be classified in two groups. One of them consists data and people and other contains technical components and both of them have dynamic nature. The users need for different types of data and this need change with the time and at the other side technology developments change the technical components of SDI and so it can be said that the concept of SDI has a dynamic and changing nature.

2.3 Data Quality

When the producer and user of data are not the same, stating the data quality will be needed and producer must describe the quality of acquired data and in the other side the user should describe requirement quality for his/her work. The issue of data quality becomes more important when we want to share data. In order to reduce expenses in the field of data collection, the data that is acquired by one organization should be useable for others. Descriptions about data quality should assure the user that the data are suitable for his/her purposes. There are some known criterions for stating quality of data. These elements are: lineage, positional accuracy, attributes accuracy, logical consistency and completeness.

Reasons of applicability forced Salge add another elements to describe data quality, semantic accuracy.Semantic accuracy describes the number of features, relationship, or attributes that have been correctly encoded in accordance with a set of feature representation rules. Related to the meaning of the “things” of the universe of discourse (the reality), semantic accuracy refers to the pertinence of the meaning of the geographical object rather than to the geometrical representation [5].

Each dataset that is collected or updated should be evaluated in the context of known quality parameters like integrity, completeness and accuracy. It is obvious that a perfect quality management is needed but this paper has been limited to define and impose some of data quality parameters in data acquisition process. The following cases are the issues that this article has considered:

1- Each object should have the right geometry type
2- All attributes about a feature and their relations must be considered
3- Values of attributes should have pertinence with defined attributes types
4- Integrity constraints should be taking into account (for example intersection of a ditch and a road is forbidden)

2.4 Web Feature Services

For controlling the first three challenges we used the data schemas that are available via the appropriate Web Feature Services (WFS). A web feature service enables the user to access integrated data that are stored in a server. Using WFS we can have operations such as:

- Querying on a database and retrieving features
- Finding the definition of any object and its whole attributes
- Adding new feature to database
- Deleting an object from database
- Updating an object in a database
- Lock features to prevent modification
-

The Web Feature Service (WFS) is easily one of the most valuable specifications of the OGC (one of the leaders in the field of standardization for SDI). It provides a generic way to access raw geographic data over the web. To the general user, this can be a wealth of information embedded in the map being viewed. Parts of the WMS (Web Mapping Services) tried to implement this functionality, but using WFS gives much more control over how to actually access that data [6].

Figure 2 shows the components necessary to serve geographic features and process transaction requests from client applications using HTTP.

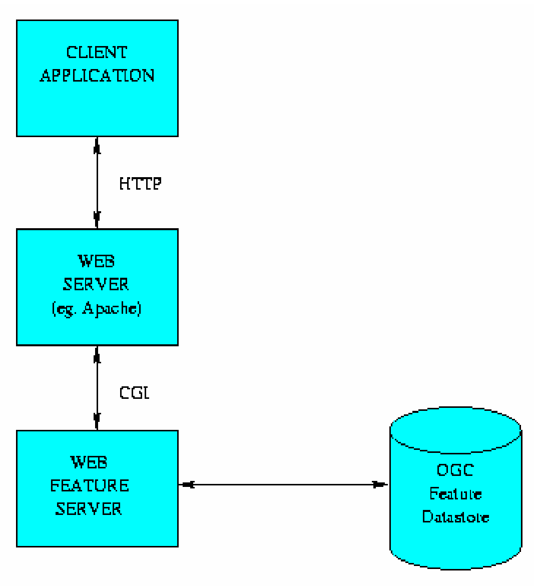

Figure2. WFS architecture

A focal objective of establishing spatial data infrastructure (SDI) is to standardize the way of producing data; SDI can provide a suitable framework for error handling purposes. Web feature services (WFS) are suitable tools for achieving this aim because they consist of information about features definitions and their complete attributes.

3. MOBILE GIS

Nowadays technologies such as: internet, wireless communication and portable devices are changing the way of using geospatial information systems from desktop to the field and on the handheld devices [7].

Devices that combine a pocket PC with a GPS receiver and cell phone and modem or other technologies, enable users to mix spatial analysis with their daily life and because of this, geospatial information systems has entered in a market with myriads of users [8].

From traditional point of view, GIS can be defined as a computer system which is used for collecting, simulating, processing, searching, analyzing and describing geographic data, in brief. GIS can be treated as an integrated computer science subject. Usually, in traditional GIS, the spacious objects and the relationship among spacious objects are unchangeable or rather static. While, under the scene of one or more moving objects in the state of moving in static object or another moving object’s scope, the spacious object and reference object (dynamic or static)'s simulating computer system is called Mobile GIS [9].

3.1 Components of Mobile Gis

Mobile geospatial information systems are the integration of various technologies like these [10]:

- Geospatial Information Systems
- Mobile software in the mould of portable devices
- Global Positioning System
- Wireless communications to access web geospatial information systems

These components can be seen elaborately and schematically in Figure 3.

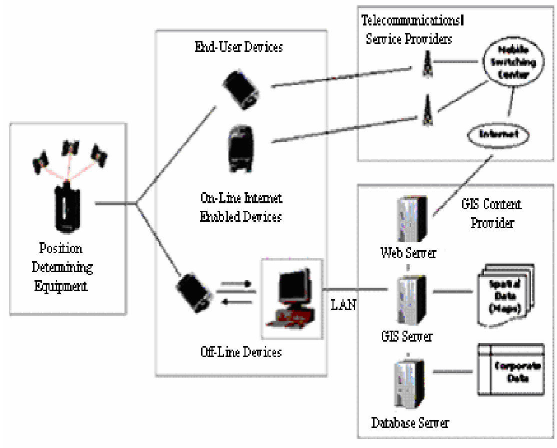

Figure3. Components of Mobile GIS

For the purpose of collecting data the required positional accuracy is high and among different techniques for positioning in mobile area, the method of A-GPS is most accurate. A-GPS relies on the GSM network providing additional information to provide integrated GPS receivers with improved coverage compared to stand-alone receivers [11]. Since stand alone GPS positioning is subject to errors (commonly classified as: ephemeris data, satellite clock, ionosphere, troposphere, multipath and receiver errors), the additional observations provided by an A-GPS system help improving the reliability and accuracy of positioning [4].

3.2 Architecture of Mobile GIS

With regarding the used hardware and software in Mobile GIS, the architecture of Mobile GIS for field works can be shown schematically as below [12].

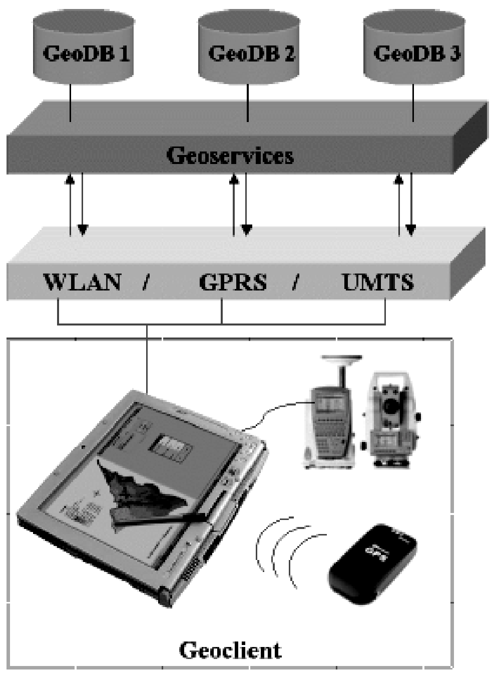

Figure4. Field based Mobile GIS architecture

The selected conventional, standardized GIS client /server interfaces like WMS and WFS 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 is 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 meeting. The usage of standardized interfaces and therewith the avoidance of proprietary developments leads to an open structure of the GIS platform.

4. FIELD DATA COLLECTION AND QUALITY ASSURANCE

Mobile data acquisition provides a new chance for the fieldwork in geosciences. 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 cannot 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 [13].

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 [14]. 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, the instruction information about how users should react to the possible errors is also defined.

4.1 Ontology based quality constraints definition

During mobile data acquisition, data quality 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, constraints have been composed with the mobile data acquisition system as an extension to the data application schema. In that conjunction 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 like "a hiking way is not allowed to be intersected with a ditch" cannot be provided by the normal GML application schemas. Therefore a way of extending the GML application schema has to be investigated. An ontology based method can be used for that. 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 [15], 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, 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 [16]. 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 3.

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, user's 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.

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 presents 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.

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.

```xml
<ruleml:imp>
    <swrlagis:constraintID rdf:datatype="http://www.w3.org/2001/XMLSchema#positiveInteger">2</swrlagis:constraintID>
    <swrlagis:severity>to apply with caution, users reaction necessary</swrlagis:severity>
    <rdfs:comment>Ways that are publicly accessible should not be closer to ditches than 10 Meters</rdfs:comment>
    <swrlagis:correctionInstruction>Set way attribute "publicly accessible" to false, close the way for public access </
swrlagis:correctionInstruction>
    <ruleml:_body>
        <swrlx:classAtom>
            <owlx:Class owlx:name="Way"/>
            <ruleml:var>way</ruleml:var>
        </swrlx:classAtom>
        <swrlx:classAtom>
            <owlx:Class owlx:name="Ditch"/>
            <ruleml:var>ditch</ruleml:var>
        </swrlx:classAtom>
        <swrlx:individualPropertyAtom  swrlx:property="publiclyAccessible">
            <ruleml:var>way</ruleml:var>
            <ruleml:var>x</ruleml:var>
        </swrlx:individualPropertyAtom>
        <swrlx:builtinAtom  swrlx:builtin="swrlb:equal">
          <ruleml:var>x</ruleml:var>
          <owlx:DataValue owlx:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</owlx:DataValue>
        </swrlx:builtinAtom>
    </ruleml:_body>
    <ruleml:_head>
        <swrlagis:builtinAtom swrlagis:builtin="dWithin">
            <ruleml:var>way</ruleml:var>
            <ruleml:var>ditch</ruleml:var>
            <swrlagis:specification>Forbidden</swrlagis:specification>
            <swrlagis:distance>10m</swrlagis:distance>
        </swrlagis:builtinAtom>
    </ruleml:_head>
</ruleml:imp>
```

Figure5. Example of quality constraints defined in SWRL

5. CONCLUSION

As we showed in this paper using WFS during field data collection is more effective way for ensuring the quality of collected data and for sharing data between different geographic information communities. Using WFS can solve problems such as shortcomings in the number of feature attributes, mistake in the value of attributes and in the geometry of features and different understandings from same features that is one of the main purposes of establishing Spatial Data Infrastructure. And for the purpose of imposing logical constraints we can use ontology languages like SWRL.

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.

REFERENCES

[1] Budic and Pinto, 1990; Rhind, 1999; Lemmens, 2001

[2] Groot, R. & McLaughlin, J. D. (2000), Geospatial Data Infrastructure: Concepts, Cases, and Good Practice, Spatial Information Systems, eds. Burrough, P. A., Goodchild, M. F., McDonnell, R. A., Switzer, P. & Worboys, M., Oxford University Press, New York.

[3] Rajabifard, A. & Williamson, I. P. (2001), 'Spatial Data Infrastructures: Concept, SDI hierarchy and future directions', Proceedings of GEOMATICS'80 Conference, Tehran, Iran, Available:
<http://www.geom.unimelb.edu.au/research/publications/IPW/4_01Raj_Iran.
pdf> (17/04/2001).

[4] Davies, Jessica. "Expanding the Spatial Data Infrastructure to Support Spatial Wireless Applications", PhD Thesis, 2003.

[5] F.Salge. "Semantic Accuracy," in S.C. Guptill and J.L. Morrison (Eds.): Elements of spatial data quality, Elsevier Science Ltd., BPC, Wheatons Ltd, Exeter, UK, 139-151, 1995

[6] WFS, Powered by Atlassian Confluence, the Enterprise Wiki. (Version: 2.5.5 Build:#811 Jul 25, 2007)

[7] Wilson, J. D., (2000), Mobile Technology takes GIS to the Field, GeoWorld, Vol. 13, Issue 6, pg. 32 to 36

[8] Hunter, Andrew. J. S. "Mobile GIS as if Field Users Mattered: Small is Ubiquitous but can Speech be Recognized?", MS.c Thesis, 2002.

[9] Li Luqun. , Li Minglu. "A Research on Development of mobile GIS architecture", Environmental Informatics Archives, Volume 2 (2004), 920-926

[10]
Http://www.gisdevelopment.net/technology/mobilemapping/ma04088.htm (1 of 3) 18-07-2007

[11] Zhao, Y. (2000), 'Mobile phone location determination and its impact on intelligent transportation systems', IEEE Transactions on Intelligent Transportation Systems, vol. 1, no. 1, March, pp. 55-64.

[12] Stephan Mäs, Wolfgang Reinhardt, "Concepts for quality assurance during mobile online data Acquisition"

[13] Fei Wang, Stephan Mäs, Wolfgang Reinhardt and Admire Kandawasvika "Ontology Based Quality Assurance for Mobile Data Acquisition"

[14] 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

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

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