Handling the Semantics of Sensor Observables within SWE Discovery Solutions

Simon Jirka, Arne Bröring, Theodor Foerster

***Abstract*— When searching for sensor data, sensor instances, or Sensor Web Enablement (SWE) services the description of the observed phenomenon plays an important role. Obviously, every user searching for sensor data needs to specify in which kind of sensor data he is interested. In current SWE applications, the information about the observed phenomenon is provided as a unique link encoded as a Uniform Resource Name (URN). However, relying on those URNs to perform string based search for sensor observables has serious drawbacks when it comes to realizing advanced sensor discovery tools as the meaning of the observables is ignored.**

This work presents an approach that makes use of semantic annotations of SWE resources. The presented solution relies on a dictionary for sensor observables, the Sensor Observable Registry (SOR). This dictionary comprises URNs identifying observables, definitions of these observables in natural language, and pointers to formal phenomenon definitions contained in ontologies. This makes it possible to rely on existing reasoning mechanisms for determining equivalent or related observables (e.g., specializations or generalizations) to the one specified by a user.

Finally, an approach is presented, how the SOR can be used for enhancing the sensor discovery process by linking it to sensor catalogues and registries.

***Index Terms*—Sensor Web Enablement, Semantic Annotations, Sensor Discovery, SWE Discovery**

I. Introduction

SENSOR networks are used in a broad variety of applications ranging from environmental monitoring and public health to disaster management [1] and monitoring of public infrastructures [2].

For enabling the flexible and interoperable integration of sensors and sensor data into any kind of application, the Sensor Web Enablement (SWE) working group of the Open Geospatial Consortium (OGC) has developed a framework of different standards. These standards specify data formats as well as web service interfaces which together form the foundation for an interoperable web based architecture that allows to access sensor data, to subscribe to alerts and to control sensors [3].

However, at the moment the task of discovering sensors, sensor data and sensor web services is still work in progress within the SWE architecture. Currently, first approaches such as the discovery framework described in [4] are under development.

A particular challenge with regard to the discovery of sensors is the need for a means to describe which phenomenon (also referred by the term “sensor observable” or just “observable”) a sensor measures or for which phenomena a user wants to find sensors. The usual approach within SWE applications is to refer to the phenomena that are observed by a sensor through a unique identifier (i.e. a URN).

However, one drawback of identifying phenomena based on strings becomes quickly apparent for sensor discovery. One important issue is that such identifiers may vary across different domains, although their meaning is the same. In consequence, every user searching for certain sensors needs to know all possible identifier strings that point to the phenomenon he is interested in. Another problem occurs when users want to search for sensors that measure something which is equivalent or similar to the phenomenon the user would like to retrieve sensor data about. Also in this case a purely string based approach would not lead to satisfying search results. In both cases the user is not able to identify the semantics of the observables and thereby might not get a satisfying result matching his search. This problem of missing semantics is also relevant for the case of automatic discovery and integration of sensor data.

This paper presents an approach how the semantics of sensor observables can be integrated into the sensor discovery process. In particular, the so-called Sensor Observable Registry (SOR) is proposed for managing the semantics of sensor observables. Furthermore, this paper shows how the proposed SOR can be coupled to sensor registries and catalogues to provide basic semantically enabled sensor discovery mechanisms.

In Section II the related work and concepts of sensors and semantic discovery are reviewed. After this, Section III gives an overview of the functional requirements for the SOR. Based on these requirements, the SOR including its functionality will be described in Section IV. Subsequently the

Manuscript received January 18, 2009. This work was supported by the following two projects co-funded by the Information Society and Media DG of the European Commission: GENESIS (project reference: 223996) and OSIRIS (contract number 0033475).

Simon Jirka is researcher at the Institute for Geoinformatics of the University of Münster, Germany (phone +49-251-83-31962; fax: +49-251-83-39763; e-mail: jirka@uni-muenster.de).

Arne Bröring is manager at the Open Source initiative 52° North, Münster, Germany (e-mail: broering@52north.org).

Theodor Foerster is postdoctoral researcher at the Institute for Geoinformatics of the University of Münster, Germany (e-mail: theodor.foerster@uni-muenster.de)

application of the SOR to enhance the sensor discovery process with catalogues and registries will be illustrated in Section V. After presenting the results of the implementation and evaluation of the SOR in Section VI, this paper will end with a conclusion including also an outlook on future research steps.

II. Background

The sensor discovery solution introduced in this paper is addressing especially the SWE architecture of the OGC. An overview of the SWE framework is described in [3].

Although the SWE architecture does not yet contain a standardized sensor registry, it offers already means for providing metadata about sensors for enabling sensor discovery: the Sensor Model Language (SensorML) [5]. As SensorML has a relatively generic character and defines most of its contents as optional, there is a need for specifying a minimum of necessary metadata to enable sensor discovery. Such a discovery profile for SensorML which defines the required metadata elements can be found in [6]. However, both of these documents do not yet address the handling of the semantics of sensor observables. Usually the observables are referred within SensorML documents by textual identifiers (e.g. URNs) which are not further linked to a formal definition of their meaning.

An approach for a sensor discovery framework based on the SWE architecture is described in [4]. Within this article the concept of the SOR has first been introduced. A more detailed version of the SOR interface as it was developed within the European OSIRIS project can be found in [7].

Research on the Semantic Sensor Web [8] investigates the role of semantic annotation, ontologies, and reasoning to improve discovery on the Sensor Web. It combines OGC's vision of a Web of sensors with the reasoning capabilities of the Semantic Web. The integration of semantics into the Sensor Web standards is not only important for discovery, but also the plug & play of sensors into the Sensor Web raises semantic challenges [9]. A method for linking geosensor databases with ontologies has been presented by Hornsby [10]. An ontological analysis of the OGC standards on observations and measurements has been done by Probst [11]. Stasch et al. [12] elaborated an algebraic formalization for the main concepts of the Sensor Web. Henson et al. [13] propose a semantically-enabled Sensor Observation Service (SemSOS). Recent approaches to enrich geospatial services with semantics include the OWL-Profile for the Catalogue Service Web (CSW) suggested by Stock et al. [14] and the development of a transparent semantic enablement for Spatial Data Infrastructures [15]. The latter approach defines specific profiles for Web Processing Service (WPS) and CSW serving functionality for reasoning and ontology look-up respectively. Those solutions offer a high degree of flexibility and generality resulting in complex specifications which are hard to implement. This work follows a more focused and simple approach which is easy to realize.

III. Requirements of the SOR

The work on the SOR described in this article was driven by a set of use cases that had to be solved in several research projects, especially the EC projects OSIRIS[1] and GENESIS[2].

A typical use case is that a user wants to query for sensors measuring a certain phenomenon. However, before submitting such a query he needs to know in which way the phenomenon he is interested in shall be specified in a search request. Within the SWE framework, usually the observables of sensors are referred by URNs. But as these URNs may vary depending on the application domain and as the observable names might even be unknown to the user, there is a need for a dictionary to look up the possible observable URNs and their associated definitions.

In other situations users need to interpret metadata about observations or sensors. In this case they usually receive SensorML encoded metadata documents. SensorML follows in general a soft-typing approach, meaning that information about the sensor, such as the input and output parameters, are identified through unique identifiers - usually URNs. As these URNs may not be self-explanatory, users need to look up the exact definition of the according observable in order to interpret the metadata correctly. This creates a need for a dictionary that is capable of resolving observable URNs and to return the associated definitions and explanations.

Often, the same phenomenon is named differently within different application areas (e.g.: hydrologists may use different names for the same phenomenon than geographers) or even within the same domain. To enable the seamless use of sensor data across multiple domains an approach is needed to handle such naming differences. With regard to enhancing sensor discovery mechanisms there is a need for a solution that is capable of determining all URNs which identify the same phenomenon. To achieve this goal, the semantics of phenomena identified by URNs need to be taken into account in a way that is transparent to the user.

Finally, when searching for sensors, users may be interested in sensors that observe a phenomenon belonging to a broader thematic domain or to a certain category of phenomena. For example, during an accident in a chemical plant a cloud of air pollutants is released. Immediately after the release of the air pollutants there is no information available which kinds of gases were released during the accident. Thus, a user would need to search for all sensor types that measure any kind of air pollutant for detecting which chemical substances are contained within the pollutant cloud. For achieving this aim, a user might rely on his personal knowledge for specifying in a search request explicitly every type of air pollutant he is aware of. This would lead to a long list of possible phenomena within a search request for sensors (e.g., searching separately for sensors measuring CO, NO2, SO2, H2CO, etc.). The problem with this approach is, however, that the process of compiling such a list might be cumbersome and time consuming. Furthermore, there is a risk of leaving out one or

[1] http://osiris-fp6.eu/
[2] http://www.genesis-fp7.eu/

more relevant substances. Alternatively, it may be easier for users to just specify one single phenomenon that is more generic so that all potential sensors are covered. In the example the user could state that he would like to find all sensors for the phenomenon “air pollution”. Based on such single term a search request may lead to a complete result, matching the user’s needs. To enable such a successful search the semantics of the phenomena measured by sensors are crucial. This leads to the requirement to determine all specializations, generalizations or equivalent concepts of a specific phenomenon.

From these use cases two main requirements were derived that comprise the core functionality of the SOR described in Section IV:

- a mechanism for making descriptions/definition of observables that are identified by a specific URN accessible in a standardized way
- an approach for determining those observables that are related to a given one (equivalency, specialization or generalization)

The next section describes how these two main requirements are addressed by the definition of the SOR interface.

IV. The Sensor Observable Registry

Based on the requirements explained in Section III the SOR is designed as a web service which provides access to observable definitions, resolves URNs, identifies observables, and exploits semantic relationships between different observables.

The SOR currently offers three operations that can be requested by a client.

The GetDefinitionURNs operation allows clients to retrieve lists of URNs of definitions that are supported by a specific SOR instance. In addition clients may specify text strings to search for specific URNs within the SOR. A paging mechanism is foreseen that allows the retrieval of subsets of long URN lists. The primary use of this functionality concerns the development of SOR clients as well as according graphical user interfaces. For example, this operation allows creating lists from which users are able to select those URNs that identify the phenomena they are interested in.

For resolving an observable URN and for retrieving the definition of the observable, clients can invoke the GetDefinition operation. This operation takes the URN that identifies a certain phenomenon as input parameter and returns an according dictionary entry.

Finally, the GetMatchingDefinitions operation provides those URNs that identify observables related to another given phenomenon in a user-specified way. In our SOR concept we require the “generalization”, “specialization” and “equivalency” relations.

Within the following subsection these three operations will be explained more detailed.

A. *Retrieving Available Observable Identifiers*

For developing comfortable user interfaces it is necessary to be able to retrieve information about the observable identifiers that are contained within a specific SOR instance. This makes it possible to offer a list of available observable identifiers from which users can select those phenomena they are interested in. The GetDefinitionURNs operation provides such a list of available observables. Optionally, a client is able to specify a text string that shall occur within the returned URNs (e.g. for requesting all URNs containing the substring “air”). Additionally, a paging mechanism is available for requesting subsets of a longer URN list. Fig. 1 shows the typical workflow when using the GetDefinitionURNs operation.

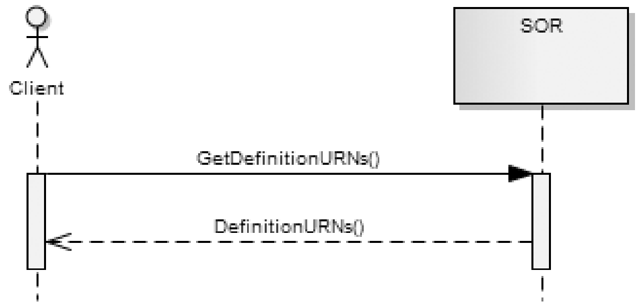

Fig. 1. Workflow for retrieving available observable identifiers

B. *Retrieving Definitions of Observables*

As explained in Section III, one of the core functionalities of the SOR is to resolve URNs identifying the definitions of phenomena that are observed by sensors.

This operation takes the identifier of such a phenomenon (in case of the SOR the identifiers are currently restricted to URNs) and returns the according dictionary entry (see Fig. 2). This dictionary entry contains the phenomenon id (URN), a short textual explanation of the observable and a link to an ontology where the semantics of the observable is formally defined.

For example a GetDefinition request for the URN “osiris:def:phenomenon:OSIRIS:Temperature” will lead to the following response:

- Phenomenon/observable id:
  osiris:def:phenomenon:OSIRIS:Temperature
- Ontology link:
  http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature
- Explanation:
  the degree of hotness or coldness of a body or environment

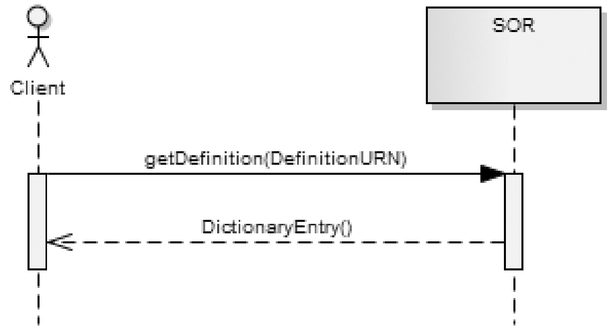

Fig. 2. Workflow for retrieving the definition of an observable

C. *Determining Related Observables*

The GetMatchingDefinitions operation allows SOR clients to retrieve a list of URNs that point to observables which are related to another given phenomenon in a certain way. Currently the SOR is able to handle the relations “generalization”, “specialization” and “equivalency”. However, for the future more complex relationships (e.g., similarity) may be included by relying on existing semantic web services.

Within a GetMatchingDefinitions request, a user is able to specify the following criteria:

- The URN of the phenomenon for which related phenomena shall be retrieved
- The type of relation between phenomena that shall be used for reasoning and finding matching phenomena (“generalization”, “specialization” and “equivalency”)
- The number of intermediate steps that are allowed in case of transitively related phenomena (e.g. air temperature in 2 meters height is a specialization of air temperature which is again a specialization of temperature)

As a result, the user receives a list of those URNs of observables matching the submitted request. For example a request for specializations of osiris:def:phenomenon:OSIRIS:Temperature with no intermediary steps could lead to the following list of URNs:

- urn:osiris:def:phenomenon:OSIRIS:AirTemperature
- urn:osiris:def:phenomenon:OSIRIS:WaterTemperature

The typical workflow when executing the GetMatchingDefinitions operation is shown in Fig. 3.

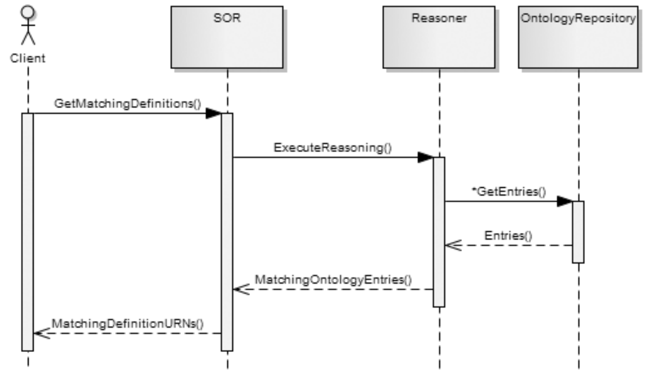

Fig. 3. Workflow for retrieving related observables

V. Enhancing Sensor Catalogues Using the SOR

The primary reason for developing the SOR was to enhance the sensor discovery process. Conventional approaches for discovering resources in web based architectures are usually based purely on string matching mechanisms whereas the meaning of the resources is not taken into account. This means that a search engine returns all resources that match to a user defined string or sub-string.

As explained in Section III this approach has several shortcomings when searching for sensors that measure a certain phenomenon.

The GetMatchingDefinitions operation plays a central role for realizing such an improved and semantically extended sensor discovery.

Fig. 4 shows how SOR functionality can be integrated into a sensor catalogue or registry. In case of conventional catalogues, the user only interacts with a sensor catalogue by querying the registry using an identifier string pointing to the phenomenon of interest. Additionally, the user is able to describe within a search request that he is also interested in generalizations, specializations or equivalent observables. Based on this information the sensor catalogue is able to connect to a SOR instance and to retrieve the identifiers of all phenomena that are related to the one specified in the search request. Consequently, the sensor catalogue queries its internal search index not only for the phenomenon identifier initially specified by the user but also for all phenomenon identifiers returned by the SOR in the GetMatchingDefinitions response (using the additionally specified relations such as generalization).

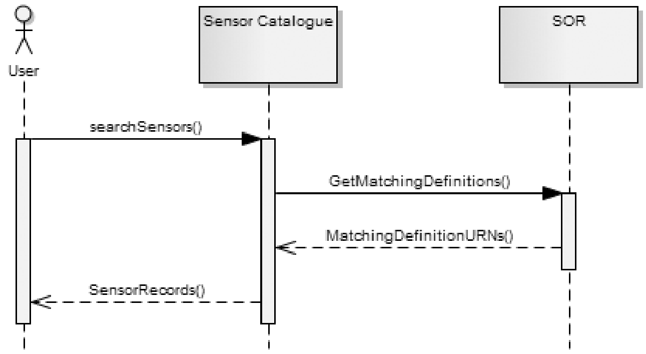

Fig. 4. Interaction between sensor catalogues and SOR

VI. Implementation and Evaluation

A first implementation of the SOR was completed within the European project OSIRIS by Thales Communications according to an earlier version of the SOR specification [7].

In context of the European project GENESIS, we have contributed an enhanced version of the SOR complying with the specification described above. This version is available as free software through the open source initiative 52° North[3]. The 52° North SOR implementation has been developed using the Java programming language and can be easily deployed within a broad range of different web service containers.

Internally this implementation relies on a XML based phenomenon dictionary as it is defined within the SWE Common specification [5]. This dictionary contains entries consisting of the URN of an observable, its natural language definition and a link to an according ontology (see Section IV.C).

Thus, the implementation of the GetDefinition observation simply relies on looking up the according observable URNs within the dictionary and then returning the according dictionary entries.

For realizing the GetMatchingDefinitions operation this dictionary is used as well: in a first step the observable URN is resolved through the dictionary to a link to an ontology which is subsequently forwarded to the reasoner.

The prototypical SOR implementation relies on the Jena Semantic Web Framework[4] as the reasoner and on the SWEET ontology[5]. This implementation showed that the approach described in this paper is working well. However, for the future the integration of additional ontology repositories will be an important work item to offer a more comprehensive set of phenomenon definitions and concepts.

Already the tests of the first SOR implementation within the OSIRIS project in conjunction with a project specific sensor registry lead to very positive results. It was noted that even quite simple relationships such as specialization, generalization and equivalency were able to improve the discovery process of sensors significantly. In a basic scenario, users were able to specify their interest in a certain phenomenon and all further phenomena that are related to it. All underlying reasoning and discovery mechanisms, however, were completely transparent to the users so that they did not need to understand the SOR.

The experiences of the OSIRIS project lead to the SOR specification as it was presented in the previous sections (the full specification is documented in [16]). The most significant enhancements comprised a broader set of parameters within a SOR request to specify explicitly in which of the relationships (generalization, specialization and equivalency) the user is interested in. Furthermore, for transitively related phenomena, a depth parameter was introduced to restrict the result set returned by the GetMatchingURNs operation to a certain vicinity of the phenomenon specified by the SOR user (in case of searching for specializations or generalizations).

The integration of the SOR into the sensor discovery process relying on OGC Catalogues is currently still ongoing work.

VII. Conclusion

This article introduces the Sensor Observable Registry to handle the semantics of sensor observables within the sensor web architecture. A special focus is put on the integration of the SOR into the sensor and SWE discovery process.

In our experiments the SOR shows that it improves the sensor discovery process significantly. It allows users to formulate more comprehensive search requests compared to purely string based approaches. Instead of searching only for exactly one string identifying an observable, the user is able to specify the concept of the phenomenon (using URNs) and to additionally query for related phenomena.

The approach presented in this article is a first solution that applies semantic concepts to the specific area of sensor observables. In particular, the SOR accesses ontologies describing sensor observables and performs semantic reasoning.

Both of these aspects are of interest to a broader community across multiple domains. Consequently, it should be investigated, if the access to ontologies as well as the reasoning based on these ontologies could be encapsulated by more generic web service interfaces.

Currently the SOR offers only relatively simple reasoning functionality as it focuses on investigating hierarchical relationships between phenomena by subsumption reasoning. This makes it possible to support the relationships generalization, specialization as well as equivalency between observables. For future extensions a much higher flexibility and degree of freedom is desirable. Consequently reasoning mechanisms beyond subsumption like non-standard inference such as finding the most specific concept, the least common subsumer as well as similarity reasoning should be included. An approach such as the semantic enablement architecture for SDIs as proposed in [17] might be a next step. This architecture introduces two new interfaces: the Web Ontology Service (WOS) and the Web Reasoning Service (WRS) which enable the standardized access to ontologies as well as to reasoning functionality. Finally, the current focus on handling

[3] http://52north.org/sensorweb
[4] http://jena.sourceforge.net/
[5] http://sweet.jpl.nasa.gov/ontology/

phenomena and their definitions should be broadened in future to incorporate the SOR into further application domains.

References

[1] S. Jirka, A. Broering, C. Stasch, “Applying OGC Sensor Web Enablement to risk monitoring and disaster management”. In *Proc. GSDI 11 World Conference*, Rotterdam, The Netherlands, 2009.

[2] E.H. Juerrens, A. Broering, S. Jirka, “A Human Sensor Web for Water Availability Monitoring”. In *Proceedings of: OneSpace 2009 - 2nd International Workshop on Blending Physical and Digital Spaces on the Internet.* Berlin, Germany, 2009.

[3] M. Botts, G. Percivall, C. Reed, C., J. Davidson, *OGC White Paper - OGC Sensor Web Enablement: Overview And High Level Architecture.* Wayland, MA, USA: Open Geospatial Consortium Inc. 2007.

[4] S. Jirka, A. Bröring, C. Stasch, “Discovery Mechanisms for the Sensor Web”. 2009, Sensors 9, no. 4: 2661-2681. Available: http://www.mdpi.com/1424-8220/9/4/2661

[5] M. Botts, A. Robin, *OpenGIS Sensor Model Language (SensorML) Implementation Specification 1.0.0*. Wayland, MA, USA: Open Geospatial Consortium Inc. 2007.

[6] S. Jirka, A. Broering (2009): *SensorML Profile for Discovery. OGC Public Engineering Report*. Wayland, MA, USA: Open Geospatial Consortium Inc. 2009.

[7] OSIRIS Consortium. *Deliverable 6001 Revision A - OSIRIS Architecture Specification and Justification*. Paris, France: OSIRIS Consortium, 2008.

[8] A. Sheth, C. Henson, S. Sahoo, “Semantic Sensor Web,” *IEEE Internet Computing*, pp. 78–83, 2008.

[9] A. Bröring, K. Janowicz, C. Stasch, W. Kuhn, “Semantic Challenges for Sensor Plug and Play,” in *Proc. Web & Wireless Geographical Information Systems (W2GIS 2009)*, Maynooth, Ireland, 2009, pp. 72–86.

[10] K. Hornsby, K. King, “Linking geosensor network data and ontologies to support transportation modeling,” in *Proc. GeoSensor Networks: Second International Conference, GSN 2006*, Boston, USA, 2008, pp. 191–209.

[11] F. Probst, “Ontological Analysis of Observations and Measurements,” in *Proc. Geographic Information Science, 4th International Conference, GIScience 2006*, Münster, 2006, pp. 304–320.

[12] C. Stasch, K. Janowicz, A. Bröring, I. Reis, W. Kuhn, “A Stimulus-Centric Algebraic Approach to Sensors and Observations,” in *Proc. 3rd International Conference on Geosensor Networks*, Oxford, UK, 2009, pp. 169–179.

[13] C. A. Henson, J. K. Pschorr, A. P. Sheth, K. Thirunarayan, in *Proc. International Symposium on Collaborative Technologies and Systems (CTS 2009)*, Baltimore, USA, 2009.

[14] K. Stock, M. Small, Y. Ou, F. Reitsma, “OGC Discussion Paper 09-010 - OWL Application Profile of CSW” Wayland, MA, USA: Open Geospatial Consortium Inc. 2009.

[15] K. Janowicz, S. Schade, A. Bröring, C. Kessler, C. Stasch, “A Transparent Semantic Enablement Layer for the Geospatial Web,” in *Proc. Terra Cognita 2009 Workshop - In Conjunction with the 8th International Semantic Web Conference*, Washington DC, USA, 2009.

[16] S. Jirka, A. Bröring, *OGC Sensor Observable Registry Discussion Paper*. Wayland, MA, USA: Open Geospatial Consortium Inc. 2009.

[17] K. Janowicz, S. Schade, A. Broering, C. Kessler, P. Maue, C. Stasch, “Semantic Enablement for Spatial Data Infrastructures” *in Transactions in GIS*, 2009, forthcoming.