**NCGIA** The National Center for Geographic Information and Analysis

# Report of the

# NCGIA Specialist Meeting on Spatial Webs

**Santa Barbara, California**
**December 2-4, 2004**

**By**

**Michael F. Goodchild, Phaedon Kyriakidis, Matt Rice, and Philipp Schneider**

**National Center for Geographic Information and Analysis, University of California, Santa Barbara**

**Sponsored by NCGIA, University of Redlands, The National Geospatial Intelligence Agency, and Esri as part of the project for**

***Strategic Enhancement of NGA's Geographic Information Science Infrastructure***

**April 2005**

# Table of Contents



## Appendices

## Executive Summary

On July 30, 2004 the National Center for Geographic Information and Analysis (NCGIA) released a Call for Participation in the first Specialist Meeting on Spatial Webs, to be held December 2-4, 2004 at the Upham Hotel in Santa Barbara, CA. The Specialist Meeting was in support of a multi-year, collaborative research program titled "Strategic Enhancement of NGA's Geographic Information Science Infrastructure", involving The NCGIA, The University of Redlands, The National Geospatial-Intelligence Agency (NGA), and Environmental Systems Research Institute (ESRI, corporate partner). This multi-year research effort combines a professionally-oriented educational program centered at the University of Redlands and a basic Geographic Information Science (GIScience) research effort centered at the NCGIA in Santa Barbara. Between August and November, 2004, the specialist meeting organizers Michael F. Goodchild and Phaedon Kyriakidis received many requests for participation and made several direct invitations to key individuals working in the areas of geospatial semantics and geospatial interoperability. By November 2004, thirty individuals had been selected to participate in the specialist meeting. The goal of the meeting organizers was to bring together researchers from academia, government, and the private sector in a three-day program of keynotes, plenary presentations, and small-group discussions. As with all NCGIA specialist meetings, its objectives were to survey the state of the art, to identify and prioritize research topics, and to build a community of interested researchers.

The specialist meeting focused on the following four sub-themes, each of which presents significant issues for interoperability and spatial webs:

- Syntactic interoperability, and the adequacy of current metadata standards;
- Semantic interoperability, and technologies for overcoming differences of meaning;
- Accuracy, and the ability of data sets of different accuracy to interoperate;
- Spatial support, and technologies for re-sampling and interpolation.

As expected, the call for participation yielded a very diverse set of participants from government, academia, and private sector, and from a variety of fields ranging from geography to engineering to computer science. The meeting participants each contributed a position paper before attending the meeting, and many participants had a chance to elaborate on relevant research while at the meeting. Several common themes emerged from the wide-ranging meeting discussion, and are presented here as a means of directing our research agenda and activities.

First, many participants expressed interest in the development of data sharing tools and standards, particularly those that allow for integration and reuse of data on the Web. Two key technological components of this effort — the Resource Description Framework (RDF) and Web Ontology Language (OWL) — are used for representing information on the Web. OWL is a vocabulary extension of RDF, and is used primarily as a "markup language for publishing and sharing ontologies on the Web" (W3C). Meeting participants are eager to see and help in the development of spatial components for OWL that will aid in the sharing, reuse, and ontological descriptions of spatial data.

A second, related theme that emerged from discussion is the evaluation of geospatial ontologies and in particular the human point of view in these ontologies as it pertains to geospatial data and geospatial services end-users. To create the most meaningful and useful ontologies, geospatial practitioners need to be able to understand and map between widely varying human spatial concepts, and take into account the differing contexts that humans encounter when using geospatial data tools and data. Geospatial ontologies must include discussions with domain experts in a top-down approach as well as with the normal, everyday users of spatial data tools and spatial data in a bottom-up approach. Participants suggested that formalizing geospatial ontologies too quickly using only domain experts might not produce the best results.

Meeting participants expressed interest in a compilation of best-use cases and best practice examples of spatial webs – information communities actively sharing geospatial data. Finding the best-use cases for spatial webs will lead to increased interest and effort into removing impediments and obstacles for sharing spatial data. A compelling domain for best-use cases for spatial webs was suggested to be emergency management, where the sharing of spatial data between communities is often a crucial early step in planning the logistical responses to a disaster or emergency event. Participants felt that the spatial webs research discussions would be focused by some well-defined and compelling examples and best-use cases.

Several participants discussed the importance of integrating tools and resources from different information communities in order to extend the capabilities of spatial webs. Specifically, there was interest in exploring the linkages between the high-performance/grid-computing community, the digital library community, and the GIScience community. Within the group of specialist meeting participants, there were individuals who actively work in the development of digital gazetteers that link the digital library with geographic information systems, and individuals that use grid computing in geographic information systems to solve large computing problems. Participants suggested that a fertile research area could be explored through the linkage of these three communities — linking the spatial and non-spatial information domains and extending the content and capabilities of the spatial web.

Meeting participants seemed eager to see approaches which tackled complex spatial webs problems through small, carefully defined, and achievable steps. Participants felt that the geospatial data sharing and interoperability fields contain many different component problems. They felt that a good approach is to identify small pieces of the problem, decide how the pieces fit together, and then solve the small problems one problem at a time.

Future collaboration between meeting participants will be facilitated by Web-based resources provided through the NCGIA researchers, as well as future specialist meetings and collaborative research proposals initiated by discussion at the specialist meeting.

## Introduction and Specialist Meeting Background

The NCGIA Specialist Meeting on Spatial Webs (December 2-4, 2004, Upham Hotel, Santa Barbara, California) was held in support of a multi-year collaborative research project titled "Strategic Enhancement of NGA's Geographic Information Science Infrastructure". This research project is a partnership between The National Center for Geographic Information and Analysis (Santa Barbara, Buffalo, and Maine), The University of Redlands, The National Geospatial-Intelligence Agency (NGA), and Environmental Systems Research Institute (ESRI, corporate partner). This research project has two principal research themes: (1) spatial webs -- geographically dispersed information communities actively sharing geospatial data, and (2) data integration. The NCGIA Specialist Meeting on Spatial Webs was designed to support the first of these two related research themes. The specialist meeting organizers were Michael F. Goodchild and Phaedon Kyriakidis. Rapporteurial and logistical support was provided by Matt Rice, Philipp Schneider, Jordan Hastings, and Qingfeng Guan. Organizational and administrative support for the specialist meeting was provided by Christian Brown.

The objectives of the specialist meeting were to discuss, explore, and synthesize ideas in four areas associated with spatial webs and interoperability:

- Syntactic interoperability, and the adequacy of current metadata standards
- Semantic interoperability, and technologies for overcoming differences of meaning
- Accuracy, and the ability of data sets of different accuracy to interoperate
- Spatial support, and technologies for resampling and interpolation

To achieve these objectives, the meeting organizers invited participants participants from academia, government, and the private sector, and brought them together in a three-day program of keynote presentations, small-group discussions, and synthesis. The specialist meeting was organized as five half-day sessions devoted to the following thematic topics:

1) Research project overview and an introduction to spatial webs
2) Semantic interoperability
3) Spatial data accuracy
4) Spatial data integration
5) Synthesis and recommendations for future work

The rationale for the meeting was to begin with the semantic interoperability topic after a preliminary introductory session, because it forms a broad conceptual basis for many of the problematic issues in spatial webs research. The other research topics (spatial data accuracy and spatial data integration), tend to have a more specific focus and hence were planned later on the schedule.

Each of the five half-day session was organized in a similar fashion, with a topical introduction by Dr. Michael Goodchild, followed by invited plenary presentations, group discussion, small breakout groups, and a final group discussion and synthesis. This specialist meeting format has been used successfully by the NCGIA for over 15 years to set research agendas and explore a multitude of topics in GIScience. In this instance, the specialist meeting was successful in prompting discussion and formulating items for a spatial webs research agenda that we will use in our multi-year research effort.

The remaining sections of this specialist meeting report include:

- A chronological summary of the specialist meeting
- Recommendations for future research
- Participants, position papers, and presentations (appendix A, B, C)
- A summary of our spatial webs research proposal and agenda (appendix D)

Participant presentations, where applicable, are linked from this document, as are supporting Web materials, publication references, and external reports.

## Chronological Summary of the Spatial Webs Specialist Meeting

The specialist meeting began informally with the arrival of guests at the Upham Hotel on Wednesday, December 1, 2004, and finished with their departure on December 4. The principal meeting sessions began early on Thursday, December 2 and continued until Noon on Saturday, December 4. The meeting schedule was designed to allow for presentations, in-depth discussion, breakout groups, and synthesis on three important spatial webs topics: 1) semantic interoperability, 2) spatial data accuracy, and 3) spatial data integration. Each of these three topics was briefly introduced by the meeting organizers, and explored in depth through two plenary speakers. After the plenary speakers, breakout groups were formed for in-depth discussion, followed by a synthesis of the topic by the entire group. The following meeting summary is organized chronologically according to the meeting schedule, focusing on specific contributions made by individual speakers and meeting participants. As such, it covers each of the meeting themes in turn, and is followed by a final section, “Recommendations for future work”, which is a product of the wrap-up summary session on Saturday morning.

### Thursday, December 2 2004

### Introductory presentations and welcome

Michael Goodchild welcomed the meeting participants and introduced them to the components of the NGA Infrastructure Research Project, specifically the basic research components being investigated at UCSB. The two principal research components are spatial webs (interoperability, distributed data sources, distributed services, etc.) and data integration (conflation, fusion, spatial support).

Scott Loomer (NGA) abbreviated a longer PowerPoint presentation by discussing two programs that he is involved in that are of potential interest to meeting participants. – the NGA NARP program and the DCI post-doctoral fellowship program. Loomer briefly described the nature of the academic research program (NARP) funded by NGA, and how this project contributes to NGA’s mission.

Karen Kemp introduced herself and described the Master’s in GIS program at the University of Redlands. NGA employees are being incorporated into the incoming student cohorts and educated in the Masters in GIS program during a 1-year long, intense program.

Individual participants introduced themselves and commented on their specific areas of interest.

## Plenary Session I: Spatial Information Theory and Semantics

### Martin Raubal & Michael Lutz

The first introductory presentation was given by Martin Raubal and Michael Lutz. The title of the presentation was "*Modeling and Managing the Semantics of Geospatial Data Services*". Raubal and Lutz clearly stated that their interest was not in semantics as the meaning of natural language or the attempt at capturing the meaning of natural language expressions, but rather with formalizing the semantics of technical symbols used in GIS and semantics associated with geospatial interoperability and geospatial web services. Raubal and Lutz presented their definition for ontology as "an explicit specification of a conceptualization", from Gruber (1993). They are interested in ontological relationships embodied in terms used to describe classes and variables, e.g., superclass, subclass, is_a, part_of, has_property. These terms attach meaning to objects and form relationships between objects and classes of objects. One interesting concept forwarded by Raubal and Lutz is the notion of reference systems for attribute data. We commonly use x, y, and z to specify position. They state that "users of geographic information should be able to refer thematic data to semantic reference systems just as they refer geometric data to spatial reference systems." Spatial reference systems are described in ISO 19112, temporal reference systems in ISO 19108, and semantic reference systems are being developed. A long-term goal of their research is to create "methods and tools to design and use semantic reference systems for grounding (e.g., "move"), projecting (e.g., roads and ferries to edges), and translating (e.g., cadastre to navigation). Small achievable steps toward this goal need to be identified and solved through experimental research. Some additional steps or considerations include annotation of data and services, translation between database schemata and services, the negotiation of meaning, capturing vagueness, and acknowledging the role of human perception. One of their concluding comments was "Ask first **what** needs to be said about space, then **how**". For project details see http://musil.uni-muenster.de. A few post-presentation comments of interest include Mark Gahegan's observation that the meaning and function of an object are largely dependent on the user's perspective, thus calling into question the ability to formulate specific meanings for object. For example, a chair for an adult is something to sit on; for a toddler, it is an object to hold on to in order to stand up.

An issue that caused some discussion was the assertion that there would be an objective method to judge the quality of a geospatial ontology. Some suggested that the one way to measure the quality of an ontology would be to perform some kind of comprehensive fact checking or error checking to judge the completeness. Yasr Bishr pointed out that the quality of an ontology is always relative to the user and how well it matches his/her view of the world. Gahegan suggested that at present, the best way to judge the quality of a geospatial ontology is through peer-review.

**David Mark**

The second introductory presentation of the day was given by David M. Mark from the National Center for Geographic Information and Analysis at the University at Buffalo. In his contribution titled “*Semantics of landscape: Providing information about the landscape via the semantic Web*” he explored the concept of semantics and how ethnophysiography relates to interoperability. After a short introduction to the theoretical background of semantics and the conceptualizations of landscape he presented four examples and case studies to illustrate his ideas.

In the first example, Mark showed how conceptual models for water bodies differ between French and English and how word-for-word translations are not possible without additional information due to those inconsistencies in the concepts. The second example explored the differences in conceptual models of terrain features such as hills in the aboriginal language of Yindjibarndi and in English. It concluded with the statement that there is a many-to-many relation between the set of words {mountain, hill, ridge, range} and the corresponding Yindjibarndi words {marnda, bargu, burbaa}. It would therefore not be possible to correctly translate those words between the two languages without knowing the size of the landform. In a third example, Mark continued to use the Yindjibarndi language to illustrate differences in ontologies, in this case for watercourses. Again, the conclusion is that direct translation between the two languages is not possible. In the last example, Mark presented a case study in which test subjects were asked to rate how well a given landform term was exemplified by a particular picture. The study was carried out in three languages: English, Spanish, and German. The conclusions from the study were that while for some images the terms appear to translate well, for other terms the ratings vary substantially across languages.

The presentation ended with the overall summary stating that two kinds of research are needed before a spatial semantic web for landscape can be implemented: Formal definitions of landscape feature type need to be compiled for the languages of interest and appropriate computational methods need to be developed for delimiting and classifying landscape features from DEMs and images.

**Breakout Groups I – Thursday PM**
The questions that were given to the breakout groups were:

- What is the state of the art?
- What research is needed?
- What community activities would be appropriate?

**Group 1**
Group members: Adams, Backe, Flewelling, Gahegan, Hastings, Kyriakidis, Mark, Shekhar, Wiegand

Group 1 started by discussing the standards for data sharing. They concluded that the there is only crude support for syntax and almost no support for semantics in the current standards (e.g., SDTS). The rules for data collection and interpretation are not used in any way. Furthermore, there are issues with governance and funding and it is questionable if better standards would be accepted and used if they were developed. In addition, achieving and maintaining standards is expensive and time-consuming.

The next point of discussion for Group 1 was the fact that multiple ontologies and standards are evolving for spatial data, examples being FAC, IHO, IMO, SWEET, etc. There is also an increasing need to map or cross-walk between different ontologies.

The ontology tools that are currently used (especially OWL) are underfit, meaning that they are simplistic, static because they are not tracking evolution, and rigid, since they have to be conflict-free. There is also no support for uncertainty information, which has been shown to be very important for geospatial data. Furthermore, the current ontology tools ignore the situation, both in creation and in the use of scientific data.

The last item of discussion for Group 1 was the social aspects of spatial webs. The group concluded that meaning is a social contract and that human intervention is still critical in both data creation and data use. In addition, the adoption of new methods will depend on the peer-support and –trust in the scientific community.

**Group 2**
Group members: Agouris, Bishr, Di, Freeston, Henk, Loomer, Raubal, Stefanidis, Yang

The discussion of Breakout Group 2 began with the question "For what exactly does the state of the art need to be found?" This led to a discussion about the definition of spatial webs. For the purposes of this discussion spatial webs were defined as communities of people doing research on spatial information and as conventions of spatial information users sharing spatial data across the Web and the corresponding protocols and standards to connect the nodes. Group 2 concluded that interoperability is one of the problems that cannot be solved by existing mainstream technologies and that other methods are needed. Another problem that has not been solved is how to describe a geospatial service. There are no standards so far for describing a geospatial service. Other questions that were asked were: How is a service defined in general? How can these services be standardized? And how do geospatial services talk to each other in a spatial webs environment? One suggested solution was metadata for data sharing and services.

Metadata creation is a difficult task, and an object-based approach could be a solution for an automated process.

Language differences and the linguistic as well the semantic context should be considered for semantic spatial webs. Automatic geo-referencing of text information was mentioned as an example of how geo-parsing and "geo-Google" could help with the development of spatial webs. The question that remains is should spatial webs be narrowed down to geo-Google?

**Group 3**

Group members: Armstrong, Cohn, Edwards, Frew, Lutz, Sen, Tsou

Group 3 began the discussion by talking about the origins of the semantic Web and the underlying concepts. In the light of spatial webs, the discussion quickly moved towards ontologies and how they can help to solve interoperability problems. It was said that there seems to be a big gap in the middle of the ontology hierarchy. While there are existing top-level ontologies as well as many low-level ontologies there seems to be a gap where the middle-level ontologies ought to be. This generates a large-scale cross-walking problem. Therefore, even though there are many ontologies, only a very few are actually connected to the higher level. The question that arises from this concept is whether we are pushing up the interoperability problem from the data level to the ontology level by looking at the non-interoperability of ontologies. Group 3 concluded that one of the most important research questions is investigating ways of linking the ontologies of the lower level to the higher level.

Although ontologies were the main focus of the beginning of the breakout session, the continuing discussion showed that there is quite some criticism of the ontology concept as well. It was said that the geospatial community will never be able to come up with very complex ontologies and therefore a perfect coverage with ontologies is unrealistic. Even if it were possible to generate the "ontologies of all ontologies" would we have solved the interoperability problem? Furthermore, some voices argued that it is easy to talk about how ontologies can solve all our problems, but that it is extremely difficult to actually devise the ontologies in the first place.

Group 3 concluded that researchable topics are mainly horizontal and vertical ontology integration, exploiting the concept of scale, and how spatial primitive types in the OWL language can provide reasoning. Better languages and representations are desirable.

## Synthesis I led by Mark Gahegan

The synthesis of the day was led by Mark Gahegan. He discussed the reason for having ontologies, what aspects of semantics can be captured, and the top-down vs. bottom-up approaches. Some limitations of top-down approaches related to OWL and languages for making inferences and supporting differences. Bottom-up (also discovery-based) approaches are based on citation indices, data mining, and use cases, and may be more useful than top-down ontologies for the following reasons cited by Gahegan (see also this relevant paper by Pike et al.):

- Often, vast resources are needed to develop ontologies, but provenance information can often be harvested for little or no cost, especially when using information portals or digital libraries.
- Users may not like to have semantic standards imposed on them.
- Bottom up semantics are used quite effectively already in e-commerce, e.g., in Amazon's recommendations: (80% of people who bought book X also bought book Y). Imagine such recommendations for GIS datasets, methods, workflows etc.
- Bottom up semantics can usually react to change, whereas top-down semantics need change to be carefully managed. (In the Amazon example above, the association values are in constant flux).
- Bottom-up semantics can be gathered through resource use, rather than requiring a separate mark-up exercise (though there is now some progress in automated semantic markup, in practice a human is still required in the loop). Think about how many resources we already have. How do we attach ontological markup to them?
- Bottom-up semantics typically capture aspects of situation or context that carry (in Gahegan’s opinion) at least as much meaning as ontologies do. Gahegan noted that his opinion is not shared by all.

Bottom-up semantics can be gathered through resource-use logging, rather than requiring a separate ontology-creation exercise, and typically capture aspects of use, situation or context by building associations between resources, people, times and places from which useful information can be derived by both humans and machines.

Gahegan suggested that both bottom up & top down semantics are essential components to the ways in which we understand the world, thus they both play a crucial role in defining and communicating meaning embedded in our e-resources. The difficult part is getting them to work together.

The following general thoughts were voiced by participants. The community approach for capturing ontologies emphasizes tools. The geospatial community has the problem of having only a few ontologies at the domain level. A lot more are needed. In addition, a method to cross-walk or map between ontologies is mandatory. The currently existing standards for spatial data are not semantically rich and capturing richer semantics is desirable. Who should ultimately create the semantic standards for geography? In terms of the existing tools we need to be able to capture and share how data are created. In addition, we need tools than can utilize this knowledge.

**Friday, December 3 2004**

## Plenary Session II – Spatial Data Accuracy

### Michael Goodchild

Michael Goodchild's presentation, titled "Accuracy, Support, and Interoperability" discussed several topics including geographic data and spatial webs, spatial interpolation, classification schemes, uncertainty and positional error, and a formulation for measurement-based GIS. A traditional view of geospatial data includes the older belief that every geographic object has a true position and correct set of attributes, both of which could be determined through repeated measurements and investment in improving the quality of the geographic database. Geospatial data, from this traditional viewpoint, are generated, held, and distributed by the appropriate government agency. A contemporary view of geospatial data holds that there are many potential sources of any particular data set, and that these sources vary in format, in terminology, in quality, in accuracy, in consistency, and in spatial support (the objects used to characterize a field – points, lines, and areas). One solution for the traditional view of geospatial data is to force all data downward to a common support structure where the differences can be sorted out. For the contemporary view, spatial webs can be used to help integrate the varying data and sort out the inconsistencies, errors, and problems. The areal interpolation problem was introduced as a fundamental issue in data integration. Positional uncertainty was presented as another fundamental issue in data integration with example data showing the possible conflation scenarios for street networks in Goleta, California. Errors of position for road networks may be related to databases with mixed lineages, where each data source has a different distortion. A coordinate-based GIS results in data where positional distortions are uncorrectable, because the innately error-prone process used to measure exact position is lost and therefore impossible to invert. Since each data source has a particular lineage (with regard to how the position of features was determined), the problem of correcting this error becomes unsolvable. A measurement-based GIS would allow for the correction of erroneous positional measurements because the errors in measurement propagate from one level in the hierarchy to the next in a predictable fashion. Correction of positional errors in a measurement-based GIS is smooth and does not result in the displacements and discontinuities commonly seen in large road network databases in coordinate-based GIS. At present, there has been an almost universal adoption of coordinate-based GIS. Measurement-based GIS would retain information about the measurement process, allowing for correction, update, and merger/integration with databases on other levels of the hierarchy.

After the presentation, a question was posed about how a measurement-based GIS would be implemented in a database setting where coordinates have to be explicitly stored? Goodchild suggested that participants look at a paper on object-oriented approaches to measurement-based GIS written by Matt Duckham (Duckham 2001). Jordan Henk recommended an inspection of ESRI's Survey Analyst software, which incorporates some of the measurement-based GIS ideas.

**Phaedon Kyriakidis**

In his presentation, Phaedon Kyriakidis from the University of California, Santa Barbara, introduced geostatistical concepts to deal with interoperability issues. He established the notion of spatial support and how covariance as a measure of spatial continuity can be used to translate between two datasets of different support. He presented a Kriging-based framework for transforming data from one spatial support to another. This framework can handle integrated measurements over arbitrary domains. It is simple because it utilizes standard geostatistical theory with only minor modifications. It is also comprehensive because it can handle alternative types of point covariance models, and consistent because it guarantees the reproduction of data at larger scales. One of the biggest advantages of the proposed framework is that it provides a measure of uncertainty, since it is based on Kriging concepts.

Kyriakidis then continued to introduce the concepts of spatial uncertainty and risk analysis and how they can be improved by using geostatistical simulation and Monte Carlo analysis. He explained how these concepts contribute to decision-making under uncertainty and how uncertainty and other risk concepts can help a decision maker.

**Breakout Groups II - Friday AM**

The questions that were given to the breakout groups were:

- What is special about spatial?
- How should spatial webs be embedded within the semantic Web?
- What do we mean by process and how important is it?

**Group 1**
Group members: Adams, Agouris, Backe, Armstrong, Cohn, Di, Edwards, Harmon, Goodchild, Hastings

The group began by discussing the meta-question "Why do we ask this question?" It was agreed that there is a spatial context to nearly everything but that many people do not realize this. This question was asked because the rest of the world should understand what the spatial data community is doing and we need to find ways to bridge the barrier that exists between the geospatial community and the rest of the world. Space is universal but hidden and many things might not appear spatial at the first look but could be spatial anyway. Nonetheless, being spatial is rarely the defining characteristic. This is most clearly visible in funding which is not driven by the term "space" but by keywords like "health" and "crime". This is because spatial information is never a field of itself; it is always part of a larger problem. The group defined three different ways of representing space, that is, first to spatialize something by adding coordinates to it, second to express space by some topological property like adjacency, and third by treating space as an infinite map. The geospatial community often treats space in the last way but the rest of the world does not. The discussion continued to move in the direction of OWL and whether it is necessary to have a spatial OWL. The general consensus was that a spatial extension of OWL is desirable since OWL is (at this point) the only choice for representing semantic information. There were also some voices who expressed a wish for having a spatially enhanced version of Google. In the further discussion the group decided that a spatial extension to OWL might not be enough and that it would be helpful to have a spatio-temporal OWL. The group concluded with the recommendation that research is needed on how to create a universal way of talking about space and time in the context of spatial webs.

**Group 2**
Group members: Flewelling, Gahegan, Henk, Tsou, Wiegand, Mark, Raubal, Sen, Bishr

At the beginning, the discussion of Group 2 was focused on how uncertainty and accuracy issues play an important role in spatial data analysis and how they can be visualized. The discussion then continued to move towards how spatial is special. It was said that there are three elements to spatial data, namely a spatial reference, attributes, and process. Also, the group agreed that all data are definitely temporal, whereas the question is if there a spatial component. Even more generally, the group said that all things that physically exist, exist in space and time. There was concern about the fact that spatial reference systems and semantic reference systems are not on the same level. Also, they asserted that it is not possible to have truly absolute coordinates.

**Group 3**
Group members: Loomer, Yang, Stefanidis, Shekhar, Lutz, Kyriakidis, Kemp, Freeston, Frew

Group 3 agreed that one issue that makes spatial special is that there is a mixture of locational types. On the one hand there is the absolute type, as for example data from GPS and imagery, and on the other hand there is the relative type, such as data from DGPS and other systems. Furthermore, spatial databases have complicated update procedures, since there are changes propagated through relative locations, multiple concurrent changes in near real-time, and a potential to override or modify constraints. There are obligations downstream to systems and consumers that may involve both "push" and "pull" technologies and progressive "check in" and "check out" of data. In order to distinguish spatial webs from the semantic Web, the group agreed that spatial webs are primarily graphical whereas the semantic Web is still largely textual. Spatial webs have maps, shapes, and cartographic symbols as well as spatial indexing of both the framework and the content. The group also discussed the issue of time and how time is not a single idea at the database level. The group concluded that relational database management systems do not support geospatial data well.

**Synthesis II led by Marc Armstrong**

It is possible to develop a process specification language for spatial processes. Before this can be done, it seems necessary to develop group concepts of space-time primitives.

## Friday, December 3 2004, PM

## Plenary Session III – Spatial Data Integration

### Isabel Cruz

Cruz's presentation, titled "Geospatial Data Integration", included a variety of data integration topics such as the role of ontologies in data integration, the role of standards in data integration, and the effects of new application areas, new technologies, and new architectures in data integration. Cruz spent some time reviewing issues raised by specific position papers (including the papers by Bishr, Frew, Mark, Cohn, Shekhar, Kuhn, and Gehagan). Cruz spent some time outlining problems related to syntactic heterogeneity, noting that data sources may use different syntax to represent data; and schematic heterogeneity, noting that documents can contain the same elements and attribute names but have completely different nested structures. Cruz noted the problem posed by semantic heterogeneity — that documents can have the same names for elements and attributes but different meanings.

A very interesting point raised by Cruz was her definition of data integration: the ability to manipulate (for example, query) data transparently across multiple heterogeneous data sources. The process of integration in this context is not so much about putting datasets together and permanently reconciling their specific differences as it is about developing methods to bridge their incompatibilities or differences so that a normal work flow can be carried out. This idea coincides directly with Goodchild's "contemporary view" that there are many potential sources of any particular data, and that these sources vary in format, in terminology, in quality, in accuracy, in consistency, and in spatial support (the objects used to characterize a field – points, lines, and areas); and that a spatial webs approach can be used to bridge their differences.

Cruz's views on ontology (including her definition) reflected those of Raubal, Lutz, and Kuhn. Cruz presented material on the architecture of the semantic Web (based on Berners-Lee) with a few diagrams. Cruz lists the roles of ontologies as 1) schema annotation, 2) a high-level view of sources, 3) support for high-level queries, 4) a way of doing declarative mediation, and 5) as a support for inference. Cruz then broke each of these roles down with examples and diagrams. There were only a few post-presentation questions, including a brief one about global ontologies, where Cruz responded with information from her presentation that outlined the ideas of an ontology-based data integration scheme from Fonseca and Egenhofer (1999).

**Peggy Agouris**

Agouris's presentation "Detecting and Modeling Change in Time-varying Imagery" contained an interesting look at change detection from imagery as a geospatial data conflation task. Agouris presented tools known as "snakes" for semi-automatic object extraction based on image processing techniques associated with image smoothness, curvature, continuity, and edge detection. Snake is a term from computer vision referring to the use of active contour models and interactive curve extraction from images. Agouris's routines have been used to do automated extraction of road networks from imagery. The first half of Agouris's presentation dealt with feature extraction for roads and other objects that are fixed. The second half of the presentation dealt with "automated motion imagery analysis solutions", or the automated identification of object trajectories from video, and content analysis of the extract objects to identify spatiotemporal activities and to support queries. A relevant application domain presented by Agouris includes hurricane path forecasting. Agouris concluded by stating that incorporating uncertainty in change detection improves conflation, and detection of position and shape change in video contributes to a better understanding of behavior and evolving events. One of the final parts of Agouris's presentation demonstrated how spatio-temporal behavior of moving objects can be assessed from video and modeled and documented in the form of a spatio-temporal helix. The spatio-temporal helix can then be used in queries and analyses. Goodchild made a post-presentation comment that technology such as that presented by Agouris should always be tested with the pathological cases (those that are the most difficult and problematic, i.e., red truck crashes into red truck). Goodchild related an anecdote about traffic monitoring experiments with tethered balloons, stating that the most important traffic monitoring cases might occur when the skies are clouded over (storms) or full or smoke (fire), making the technology useless in situations where it is most needed. Gahegan stated that an important aspect of Agouris's approach to change detection is to be able to characterize the nature of the changes and develop the ability to discern one set of changes from another.

## Recommendations for future work

### Saturday, December 4 2004, AM

The final plenary session was an extended opportunity for participants to recommend future research direction, prioritize goals, and plan for the future. Rather than summarize each individual comment as recorded in the raw notes, we mention a few individual comments and present a brief synthesis of each recommended topic for future research. These 5 topics for future research emerged during this final plenary session. We plan to capture and elaborate on each of these emerging themes on our project web pages and through contributions from specialist meeting participants.

The final plenary session was set up with the following questions posed by Dr. Goodchild:

What research is needed?
Where are the low-hanging fruit?
Where is the miracle?
What is needed in practice?
What can be achieved in 2 years? In 5 years?
What are the use cases?

Each participant was given the opportunity to answer this question. Some topics of continuity throughout the participant comments included:

### 1) OWL, RDF, and the development of spatial components for the OWL

The Resource Description Framework (RDF) and Web Ontology Language (OWL) — are used for representing information on the Web. OWL is a vocabulary extension of RDF, and is used primarily as a "markup language for publishing and sharing ontologies on the Web" (W3C). Meeting participants discussed the use of OWL, describing it as the primary tool (some referred to it as the 'only tool') for incorporating semantic information into data. Most participants were eager to see spatial extensions developed for OWL or modifications made to OWL so that it supports geospatial data. From the W3C Website, it appears as if OWL is being extended in several directions, e.g, OWL-S for Web Services, so this desire for a spatial extension seems to be a possibility. Participants with OWL experience include (but are not limited to) Yasr Bishr, Mark Gahegan, and Dan Adams. Dan Adams (NGA) recommended a few organizations and communities that are active in the development of the semantic Web, namely IBM, BPEL, and the European Web Standards Group. A few other groups directly involved in the development and extension of OWL include the Stanford University Knowledge Systems Laboratory, SRI International, France Telecom, Maryland Information and Network Dynamics Lab at the University of Maryland, National Institute of Standards and Technology (NIST), Network Inference, Nokia, Toshiba Corporation, and University of Southampton. Mark Gahegan noted http://www.ecoinformatics.org as the web site for a group formalizing semantics for ecological processes. David Mark noted that semantic Web development has a 'natural selection' process of acceptance of the best ontologies.

### 2) Spatial webs best-use cases and emergency management as a compelling application domain for spatial webs and data integration research

Ming Tsou related some stories from the San Diego wildfire incident during which he maintained a prominent online GIS map of damaged residential areas. Mike Freestone also spoke about emergency management and the Federal Emergency Management Agency (FEMA). Both Raubal and Lutz stressed the importance of finding excellent use cases, scenarios, and application domains, and both suggested that emergency management / emergency services was a good domain. Lutz suggested that developing good use cases and good problems is a research issue in itself, and that unless the problems are focused, the discussion "goes out the window". Shekhar suggested collecting the best practices in spatial webs and data integration, and including a thesaurus and encyclopedia of GIS. In the same vein, Jim Frew suggested that we need a 'spatial webs manifesto' for all of us to sign, i.e., a way of identifying and defining spatial webs and putting forth a spatial web agenda.

### 3) Integration of tools and linking information communities together

Mike Freeston spoke up two or three times during the final plenary session about his view for the future of digital libraries, GIS, and The Grid. He suggests that in the future, all three groups will be integrated to connect resources together, i.e., textual information meets contextual/geographic setting information meets computational resources. Later Freeston described this integration as linking the spatial/geo-referencing world and the non-spatial digital library world. Part of this effort will be accomplished by advancing the gazetteer and extending its usefulness as a tool for linking the spatial and non-spatial worlds. A related concept, referred to as 'geo-Google' was discussed at length in a few of the breakout sessions.

### 4) Ontology: evaluating ontology and the value of the human point of view in ontology

The main proponents of this theme included the Muenster Group, David Mark, and Mark Gahegan. Raubal mentioned that many groups are busy creating ontologies in various subject domains and stated that these groups need to include the perspectives of the people *using* the ontologies. Cognitive specialists (those trained in the ways that humans learn) need to be included in the discussion, because they have training in determining what is important to formalize into an ontology. Raubal also suggested that it is important to create mappings between human spatial concepts, and that formalizing different contexts is also important.

Adams suggested http://www.semwebcentral.org as a place for exchanging ontologies. He described this site as a split off from a DARPA-funded group working in the same domain.

Gahegan suggested that the first step in solving the big spatial webs and data integration problems is to “bandy about ideas”, and discuss informally how people think about and formalize domains. Capturing ontologies by asking only domain experts sometimes doesn’t work. He suggested that perhaps ontologies need to be captured informally and then formalized at a later time – contrasting top down vs. bottom up approaches to ontology.

**5) Taking small steps toward solving big problems**

Prompted by the final plenary session set-up questions, participants advocated defining and working on small achievable steps in conquering the larger spatial webs problems. Lutz asserted that the spatial webs problem is really a collection of various component problems, and a good approach would be to identify small pieces of the problem first and then decide how they fit together. A good role for the NCGIA in spatial webs research would be to help define and articulate what these small component problems are and how they fit together.

# Appendix A

## Specialist Meeting Participants

**Dan Adams**
National Geospatial-Intelligence Agency

**Peggy Agouris**
University of Maine

**Marc Armstrong**
University of Iowa

**Kevin Backe**

US Army Topographic Engineering Center

**Yaser Bishr**
Image Matters

**Tony Cohn**
University of Leeds

**Isabel Cruz**

University of Illinois at Chicago

**LiPing Di**
George Mason University

**Dan Edwards**
US Army Topographic Engineering Center

**Doug Flewelling**
University at Buffalo

**James Frew**
UC Santa Barbara

**Mike Freeston**
UC Santa Barbara

**Mark Gahegan**
Pennsylvania State University

**Russ Harmon**
US Army Research Office

**Jordan Henk**
University of Redlands

**Karen Kemp**
University of Redlands

**Werner Kuhn**
Muenster University, Germany

**Scott Loomer**
National Geospatial-Intelligence Agency

**Michael Lutz**
University of Muenster

**David Mark**
University at Buffalo

**Martin Raubal**
University of Muenster

**Sumit Sen**
University of Muenster

**Shashi Shekhar**
University of Minnesota

**Anthony Stefanidis**
University of Maine

**Ming Tsou**
San Diego State University

**Nancy Wiegand**
University of Wisconsin Madison

**Chaowei Yang**
George Mason University

# Appendix B

## Participant Position Papers

| | |
|---|---|
| Peggy Agouris | *Accuracy-aware Change Detection in Spatial Webs* |
| Marc Armstrong & Sheowen Yang | *GeoMiddleware to Support Interoperability for Grid Computing* |
| Kevin Backe | *U.S. Army's S&T on Spatial Data Integration* |
| Yaser Bishr | *What is Your Context?* |
| Tony Cohn | *SpatialWebs: PositionPaper* |
| Isabel Cruz | *Ontology Alignment for the Semantic Integration of Heterogeneous Geospatial Data Sets* |
| LiPing Di | *Geospatial Semantic Web Research at LAITS* |
| Dan Edwards | *U.S. Army's S&T on Spatial Data Integration* |
| James Frew | *Earth System Science and Spatial Webs* |
| Mark Gahegan | *A Situated Representation of GIS Resources and GIS Users* |
| Werner Kuhn | *Semantics of What?* |
| Michael Lutz & Eva Klien | *Overcoming Differences of Meaning during the Discovery and Retrieval of Geospatial Information in Spatial Data Infrastructures* |
| David Mark | *Semantics of Landscape: Providing Information About the Landscape via the Semantic Web* |
| Martin Raubal | *Cognitive Semantic Interoperability for Spatial (Web) Services* |
| Sumit Sen | *Semantics of Spatial verbs and their Role in Interoperability* |
| Shashi Shekhar | *Spatial Data Mining and Geo-spatial Interoperability* |
| Anthony Stefanidis | *Geosensor Networks and Spatial Webs* |
| Ming Tsou | *Developing Intelligent Software Agents for Semantic Interoperability on Internet GIS* |
| Nancy Wiegand | *Position Paper for the Specialist Meeting on Spatial Webs* |
| Chaowei Yang | *Syntactic Interoperability and Spatial Support through practical development"* |

# ACCURACY-AWARE CHANGE DETECTION IN SPATIAL WEBS

*A Position Paper*

**Peggy Agouris**

Dept. of Spatial Information Science & Engineering
National Center for Geographic Information and Analysis
University of Maine
Orono, ME 04469-5711

Change detection is a fundamental operation in Spatial Webs, as it is essential for geospatial database updating, conflation, and redundancy management. Considering image-based change detection in particular, current practices typically use complex algorithms (e.g., snakes, template matching) to identify and automatically extract objects in a new image. Following object extraction, a GIS database is updated by comparing the newly extracted outline to the stored prior information. Usually, if the new outline is different, the older information is replaced by it, and the database is thus updated. In this context, information is treated as deterministic in nature: any difference between two outlines is considered as change. When the two datasets compared differ in their properties (e.g., resolution, scale, accuracy) this process would result into frequent false positives: change would be detected even though the object under examination has actually remained unchanged. Given the diverse nature of datasets in Spatial Webs, this deterministic approach to change detection is clearly inadequate, forcing us to consider accuracy-aware approaches.

Addressing this problem from an image analysis point of view, we have developed a novel automated approach that integrates object extraction and change detection into a single process by using accuracy information. Our approach is meant to function within an integrated geospatial environment, whereby image analysis proceeds by having access to pre-existing information for the processed area. We assume a process where incoming imagery is analyzed to update the GIS of an area of interest. Within this modus operandi we have available from the GIS database shape information in the form of outlines and corresponding accuracy estimates for objects in the area of interest. This prior information may have resulted from image analysis (exploiting older imagery), or by any of the other established methods of collecting GIS information (e.g. traditional surveying processes).

Our overall approach is described in Fig. 1. As previously mentioned, we make use of pre-existing information for the object outline. The older outline (e.g., available from an older map or GIS layer) is projected onto the new image, using standard orientation parameters and relevant transformations. Once projected onto the new image, the older outline together with its accuracy information becomes the input for our differential snakes tool that performs change detection and versioning.

In this stochastic approach, change is detected if and only if the content of the new image (pixel values and their distribution) indicates that an object's outline lies beyond the statistical range of its last record.

Otherwise, we cannot support the hypothesis that the object has changed, and consequently, we have to limit our analysis to versioning. Versioning occurs when we can detect the outline in the new imagery with better accuracy than the one with which it was recorded in the database, even though the outline itself is not beyond the statistical range of its prior record. The results of change detection and versioning are used to update the GIS record of an object by modifying its outline and/or updating its uncertainty estimates. An outline of this updating process is shown in Fig. 1. This approach falls within the framework of accuracy-aware conflation of geospatial information to support change detection of linear objects (e.g., roads) [Agouris et al., 2001], as well as closed-curve ones (e.g., lakes) [Gyftakis et al., 2004].

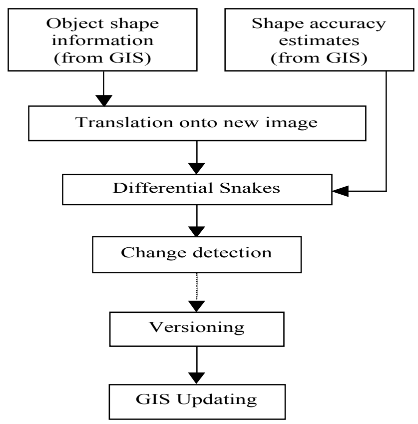


**Figure 1:** Accuracy-aware image-based GIS updating

This approach brings forward some interesting issues related to accuracy modeling and handling in Spatial Webs, including:

- *Modeling object accuracy*: Accuracy information may be available as an overall estimate for a class of features (e.g., defined by resolution and scale of available information), or it may refer to a specific object only (e.g., recorded when extracting a specific road segment from prior imagery). Thus, it could either be constant across a feature (e.g., a road known with an accuracy of 2 meters), or we may have different values for various nodes along this feature (allowing e.g., a corner partially covered by canopy to have poorer accuracy metrics than other visible nodes).
- *Storing and managing accuracy information in Spatial Web collections*: Accuracy can be stored as file metadata, or object metadata. Considering hierarchical grid architectures, we are also provided with several choices regarding the management of accuracy information in Spatial Webs.

## REFERENCES

Agouris P., A. Stefanidis & S. Gyftakis, 2001. Differential Snakes for Change Detection in Road Segments, *Photogrammetric Engineering & Remote Sensing*, Vol. 67, No. 12, pp. 1391-1399.

S. Gyftakis, P. Agouris & A. Stefanidis, 2004. Tracking Deformable Objects in Geospatial Applications, *IEEE International Conference on Image Processing (ICIP),* Singapore (in press).

# GeoMiddleware to Support Interoperability for Grid Computing

Marc P. Armstrong
Department of Geography and Program in Applied Mathematical and Computational Sciences
The University of Iowa

Shaowen Wang
Research Services of Information Technology Services, and Department of Geography
The University of Iowa

## Introduction

Middleware is software that is designed to support the interoperability of computer applications that use (and/or produce) different types of information. Middleware for geographic applications is particularly important because required information may have scale dependent relationships, which may cascade to cause problems related to level of generalization, dimension change, and categorical precision. The challenge of supporting interoperability is especially vexing when multiple types of data (from various sources and with variable error characteristics) must be meaningfully integrated for use in distributed applications. In this position paper we set out a general geographic information processing problem as a way to motivate discussion about the use of middleware to support analyses in distributed, heterogeneous Grid computing environments.

## Grid Computing and Middleware

The Grid refers to an infrastructure that enables the integrated, collaborative, and coordinated use of distributed heterogeneous computing resources, such as computers, networks, databases, and scientific instruments owned and managed by multiple organizations (Foster and Kesselman, 1999). Since the very early stages of the evolution of the Grid, middleware has been a primary focus of software development and research effort (Foster *et al.*, 2001). Middleware supports applications in distributed computing environments by providing services that enable the interconnectivity and interoperability of applications, systems and devices.

## GeoMiddleware

Several disciplines have turned to Grid computing to find solutions to computationally intensive problems; this has required the development of domain-specific middleware (see GEON[1]). Such middleware exploits the characteristics of problems and helps application scientists manage and use the Grid effectively. Geographic information science (GIScience) must develop middleware that captures important geographic characteristics of problems. This GIScience-specific Grid middleware may provide interoperable geographic analysis services that are able to reconcile different data and metadata regarding formats and semantics. This "GeoMiddleware" helps GIScientists locate, allocate, and use Grid resources effectively and efficiently (Wang *et al.*, 2002).

## A Motivating Problem

The use of grid computing to explore computationally complex network optimization problems by multiple members of a decision making group requires a considerable amount of data integration, multi-user interaction, computational decomposition and reconstitution, and

[1] http://www.geongrid.org/

distributed visualization. Each step of this complex chain of events requires access to metadata that can be used by middleware to orchestrate activities in an appropriate sequence to achieve a goal of multi-participant interaction with the problem-solving process. To accomplish this goal, data must be assembled to describe:

- A street network, possibly from different sources and with different scales; this may cause problems with distances calculations (Armstrong and Dalziel, 1989);
- Distributed demand for services can be estimated from census tabulations for small areas that are assigned to locations in the network, also introducing errors.
- Supply of existing services (if any);
- Number, type, and capacity of needed services;
- Criteria used by one or more decision-makers;
- Metadata and data that describes solutions and their characteristics;
- Maps, graphs and statistical information and metadata about individual solutions and collections of them;
- The set of available computational resources and their current status;
- Partitions of problems that can be allocated to available resources according to a scheduling plan;
- Reassembled problem sub-solutions from the partitions;
- Preferences for visualization options; and
- Information about choices, or modifications to criteria and solutions generated.

**Research Challenges**

Though a markup language, such as GML, can be used to describe some of the semantic characteristics of needed data, the use of disparately sourced data in analyses remains a problem if it is used in a distributed, heterogeneous environment. Data also may have markedly different granularities, ranging from a parameter (maybe a byte) to gigabytes of road network information. It is difficult to rationally assemble data for analysis when it is derived from different sources and with different geographical error characteristics that arise as a consequence of different compilation scales and collection purposes. For example, if road networks are used to calculate distances between demand and supply locations, differences in the level of generalization in the data for a study area can significantly affect the results obtained. As data are generalized, categories may get collapsed as well. The interaction between generalization in the geometric domain and the level of categories maintained is not always well-specified and loss of meaning can occur.

When road network data are used to represent large numbers of supply and demand locations in optimization, problems can quickly become computationally intensive. Interoperability problems may be further exacerbated by the imposition of additional (parallel) Grid computing procedures that are put into place to improve computing performance. Oftentimes, these procedures include strategies such as domain decomposition and task scheduling. Subsequent to the computation of results, additional procedures must be specified to integrate and represent solutions. GeoMiddleware must be designed to manage the integration of computational procedures as well as data fusion involving multiple sources and scales with different characteristics of errors and levels of generalization. GeoMiddleware must be able to reassemble partitions and handle multiple types of interactions of distributed users. As a consequence of these and other problems, the development of GeoMiddleware to support interoperable Grid applications is a difficult challenge.

## References

Armstrong, M.P. and Dalziel, B. 1989. Digitizing and distance estimation errors in spatial analysis: scale and network configuration effects. In: *Proceedings of the International Geographic Information Systems (IGIS) Symposium '89*. Washington, DC: Association of American Geographers, pp. 75-84.

Armstrong, M.P., Cowles, M.K., Wang, S., 2005. Using a computational grid for geographic information analysis. Forthcoming in *The Professional Geographer*.

Foster, I., Kesselman, C., and Tuecke, S., 2001. The anatomy of the grid: enabling scalable virtual organizations. *International Journal Supercomputer Applications*, **15** (3), available at: http://www.globus.org/research/papers.html, accessed on October 1, 2004.

Foster, I., and Kesselman C., 1999. *The Grid: Blue Print for a New Computing Infrastructure*. Morgan Kaufmann Publishers, Inc. San Francisco, CA.

Wang, S., Armstrong, M. P., and Bennett D. A. 2002. Conceptual basics of middleware design to support Grid computing of geographic information. In: *Proceedings of 2nd International Conference on Geographic Information Science*, September, 25 - 28, Boulder, CO, USA, 197-200.

Wang, S. and Armstrong, M.P. 2003. A quadtree approach to domain decomposition for spatial interpolation in grid computing environments. *Parallel Computing,* **29** (10): 1481-1504.

# U.S. Army's S&T on Spatial Data Integration

Kevin Backe/Technical Director U.S. Army Engineer R&D Center's Topographic Engineering Center
Daniel Edwards/Program Manager Geospatial Information Integration and Generation Tools (GIIGT)

U.S. geospatial intelligence community has moved over the past 10+ years from a "MONO- to MULTI-" environment. This "multi-" environment includes multi-sensors, multi-sources, multi-producers, multi-representation, and multi-answer. However, tools to support this transition have been lagging behind in development. There are few robust automated tools to integrate and validate Features, Imagery, and Elevation, and other spatial-temporal referenced data sources.

Traditionally, the U.S. DOD and Intelligence geospatial communities have relied on space-based sensor platforms that support the generation of relatively static geospatial data. These communities are now moving toward continuously refining and improving geospatial databases by exploiting tactical sensors, air and ground-based sensors, and humans in the field. However, the current focus of a proliferation of sensors that will provide a plethora of raw sensor feeds is not supported by a complementary level of effort to integrate/fuse all these sensors, analyze this data, and provide the commander/soldier with actionable information about the battlespace environment where he/she will operate.

To address this shortfall the U.S. Army Engineer Research and Development Center's Topographic Engineering Center has initiated an applied research program in FY03 entitled: Geospatial Information Integration and Generation Tools (GIIGT). The objective of the R&D program is to deliver tools to integrate, manage, and exploit multi-source imagery, features, and elevation data in order to present only the 'best set' of relevant geospatial information. GIIGT has established an overall guiding philosophy that applies to differing types and representations of spatial information and is the following:

1. Multi-source information is an emerging reality
2. There is no perfect sensor – positionally or informationally
3. Assume the data or derived data are partial answers – maybe right, maybe wrong
4. Store, correlate, examine and improve data in a continuous update cycle i.e.,

-Store partial answers in object database
-Examine, reinforce, refute, and mine
    -Assemble and disseminate coherent answers
    -Improve the original data

5.Generate Additional Data (as required)

This guiding view governs our research and development in the following technical areas:

<u>Geospatial Feature Fusion (Feature Linking)</u>
Feature Linking is developing software tools that will automatically match and link vector feature representations across multiple data sets to provide the user with the "best spatial and attribute information" available from multiple sources.

<u>Spatial Data Mining</u>
Spatial Data Mining is developing software tools which learn about our spatial physical world through the non-trivial extraction of implicit, previously unknown, and potentially useful knowledge from widely

differing sources of data and then extract this knowledge in such a way that it may be employed in areas such as spatial prediction, forecasting, estimation, and decision support. These approaches must use specific spatial information such as x-y coordinates and/or elevation in addition to semantic information.

Cross-Sensor Image Registration
Different sensors and sensor types have different spatial accuracies, sensor acquisition parameters, and phenomenology. The objective of cross-sensor registration is to develop a software capability to automatically correct for these absolute and relative spatial inaccuracies so that image positions can be accurately and quickly geolocated to a common uniform reference.

Digital Elevation Model Fusion
Digital Elevation Model Fusion is developing automated techniques that fuse elevation data from diverse sources into a single, coherent digital elevation model. Fusion issues being addressed involve: hole filling, adjacent/overlapping DEMs, differing quality & resolution DEMs, bare earth versus reflective surface issues.

Spatial Database Management
Efforts are underway to examine issues of storing and accessing very large spatial databases and develop object-based spatial data models.

Feature Extraction Algorithms
Feature extraction methods are combining statistical pixel classification with learning, rule-based, and geometrical relationship to generate more robust features. Further work is focusing on the direct or indirect collection of features from GPS-enabled devices in the field.

The GIIGT program is supported by contracts, small business innovative research (SBIR) initiatives, and in-house research.

From a broader perspective the U.S. Army's Terrestrial Science Basic Research Program has recently defined a geographic information science research focus area that addresses fusion of data and information at a higher cognitive level to support understanding of the impact of the battlespace environment on military operation. This geospatial research focuses on increasing knowledge about the terrain effects on modern warfare through cognitive understanding of map/digital terrain information, analysis/reasoning, modeling of geospatial data, and knowledge discovery from spatially and temporally referenced data. Relationships within geospatial data are extremely complex because it usually exhibits strong temporal and spatial correlations. Understanding shared spatial complexity within the military operating environment demands a comprehensive problem-solving approach that facilitates correlation of information across spatial and temporal scales, across multiple levels of organization, and across technology areas. New concepts and methods for acquiring and representing geospatial information can be combined with Battlespace Environment (BE) related cognitive and behavioral processes to advance fundamental understanding of the spatial dimensions of human (warfighter) and social (military echelon) dynamics. This research seeks to answer the following questions such as: How is shared, spatial information cognitively fused and assimilated for maximum understanding by the user? How does shared spatial information move within and across levels of physical and social systems?; How do humans influence shared spatial information and respond to complexity in BE systems?; What new knowledge can be gained from scientific examination of correlation, patterns, and relationships in spatial information?; and What knowledge and relationships can be found in spatio-temporal data (e.g. cycles, trends) that might be exploited in predicting the future state of the battlespace.

The payoff for this research and development on geospatial data integration and management in concert with other related efforts within DOD, industry, and universities is the ability to provide superior battlespace environment awareness, assured mobility and survivability by providing the commander/soldier/decision-maker with the "best set" of relevant geospatial data. This ability to provide the best geospatial information contributes to providing ground commanders/soldier with a superior understanding of the environment as she/he responds to a disaster/crisis at home and plans for and responds to missions in foreign lands.

# What Is Your Context?

Position Paper
Yaser Bishr, Phd
CTO, Image Matters LLC

## STATEMENT OF POSITION

Is the statement: "All Birds Fly" always true? Penguins are birds but can't fly. This statement, however, might be entirely true if its context was "In Brazil". Here we have added context to the statement, and the context being "In Brazil".

The geospatial community is busy catching up with the Semantic hype. One thing they agree on is the use of OWL as the lingua franca to develop ontology. They, however, ignored the need for an ontological model that account for Context. In the last twenty years, the notion of context has become more and more central in the theories of knowledge representation in Artificial Intelligence, cognitive psychology and linguistics, but to the best of our knowledge not in the geospatial domain.

Several domains have already elaborated their own working definition of context. In human-machine interaction, a context is a set of information that could be used to define and interpret a situation in which agents interact. In the context-aware applications community, the context is composed of a set of information for characterizing the situation in which humans interact with applications and the immediate environment [Dey, 1998]. In artificial intelligence, the context is what does not intervene directly in a problem solving but constrains it [Brézillon, 1999a]. Our working definition of context is that it is a collection of relevant conditions and surroundings that make a situation unique and comprehensible.

Our position is that Semantic Interoperability can only be achieved with general and unifying theory of context for geospatial applications. By unifying theory we mean a theory that augments ontology of the geospatial domain with context to enable agents to conduct meaningful inference and hence semantic interoperability.

## A FOUNDATION FOR A THEORY OF CONTEXT FOR GEOSPATIAL APPLICATIONS

Ontologies are shared models of a domain that encode a view which is common to different communities. Context is a model that cast a local view of shared models, i.e., shared ontologies. Context can be considered as a filter that helps scope this subset of an ontology that is relevant to a given situation. Developing a theory of context for the geospatial domains must satisfy the following requirements:

1) Context should allow a simpler formalization of axioms by defining the set of known conditions that are common to the stated axioms.

2) Context should allow us to restrict the vocabulary and the facts that are used to solve a problem on a given occasion. This requirement will enable us to scope large ontologies to those subsets that are relevant to the problem at hand.

3) The truth values of facts should be dealt with as dependent on a collection of assumptions which implicitly define context.

4) There are no absolute, context independent facts, namely each fact must be stated into an appropriate context.

5) Reasoning across different contexts should be modeled. This will enable mapping between ontologies (in context), and hence semantic interoperability.

6) A theory of geospatial context must consider the role of time, location, and other spatio-temporal aspects in determining the truth value of a given set of axioms.

McCarthy defined a context as a generalization of a collection of assumptions. The basic relation of context is *ist(c,p)*. It asserts that the proposition p is true in the context c, where c is meant to capture all that is not explicit in p that is required to make p a meaningful statement representing what it is intended to state. We go further and state that the truth value of any proposition cannot be established without establishing its context c. Context c is a collection of propositions p`, which in turn can only be true in context c`. Formulas ist(c,p) are always asserted within a context, i.e., something like ist(c', ist(c,p)): c': ist (c, p). This is inline with Russel's introduction of context, where he introduced contextual descriptions that have no meaning of their own. However every sentence in which they occur has a meaning [Arbab, 1992]. The consequences are:

- A context is always relative to another context,
- Contexts have an infinite dimension
- Contexts cannot be described completely,
- When several contexts occur in a discussion, there is a common context above all of them into which all terms and predicates can be lifted.
- There are relations between contexts. For example a context c may subsumes context c`. This enables us to propagate or lift assertions of one context into another using some predefined lifting rules.

## CONCLUSIONS

Semantic interoperability can only be achieved if we introduce context into our ontological models. To the best of my knowledge there is a lack of body of research that investigates the role of context in the ongoing activities to build ontological model for geospatial applications. This position paper intends to mobilize the research community to build a coherent general theory of context for the geospatial domain and begin to lay the foundation for such theory.

## Selected References:

[Brézillon, 1999a] Brézillon P., "Context in problem solving: A survey", *The Knowledge Engineering Review*, 1999, 14(1), 1-34.

[Dey, 1998] Dey A.K. and Abowd G.D., "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications", 1998.

**Arbab B.** (1992) "A formal language for representation of knowledge", Proceedings of the AAAI'92 Workshop on Propositional Knowledge Representation, Stanford, CA, pp. 1-8.

# Spatial Webs: Position Paper

Anthony G Cohn
School of Computing, University of Leeds, LS2 9JT, UK

]

I have a number of interests related to the topic of the workshop and I describe these briefly below. First I describe my general interest in spatial ontology, then I describe a new project on spatial data integration, and finally I discuss how work I have been conducting in cognitive vision may be relevant.

## 1. Spatial Ontology

Fundamental to any Spatial Information System, including a Spatial Web, is a characterisation of spatial entities and the relationships between them. This is an area I have worked in since the late 1980s, concentrating in particular on Qualitative Spatial and Spatio-temporal reasoning, particularly mereotopological relations over regions [5], but also on orientation [12]. The Leeds work on mereotopology is based on a fundamental notion of two entities being *connected*, and while the resulting calculus (called the Region Connection Calculus (RCC) has much in common with the 4 and 9-intersection calculi (e.g. [9]), the setting within a full first order language means that it is fundamentally more expressive, rather than simply being a *constraint language*, although for computational reasons, the reduced expressive power of a constraint language can be advantage. We have also developed a more expressive language, *Region Based Geometry* [3]. This gives a very expressive language, allowing complex shape information to be specified, but with a corresponding decrease in computational efficiency, as would be expected. An extension of RCC has been formulated to explicitly handle regions with indeterminate boundaries [6]. A review of the whole area of qualitative spatial reasoning can be found in [8].

Since many applications also have to consider how spatial entities change, and change their relationships to each other over time, we have also considered spatio-temporal calculi, in particular calculi based on spatio-temporal regions [7]. Fundamental to the notion of most notions of change is the idea that changes must, in some sense, be continuous. We have analysed a variety of notions of qualitative continuity, of varying strengths, including purely spatial, and purely temporal continuity [7].

## 2. Mapping the Underworld

Every year, in excess of four million holes are dug in UK roads to repair leaks, provide connecting services to new premises and to lay new cables and pipes[4]. Although recently installed assets may have been well mapped, location data on older services can be very poor, in some cases even non existent (except perhaps knowing the location of the terminating points). Some of the holes are unnecessary (dug in the wrong place owing to insufficient or wrong data), some cause third party damage to other underground services (or even first party damage!). More importantly, there are also considerable indirect costs owing to disruption on the roads caused by works, waste, and pollution. A recent Engineering and Physical Sciences Funding Council (EPSRC) initiative initiative is funding £1M of research to ameliorate the situation, ranging from better sensing devices to improved construction techniques to engendering better cultural relations and education of the relevant parties. The core of the problem is that there is at present insufficient and inadequate knowledge about *what is where*. The University of Leeds has received funding under this initiative to tackle this central issue. Even with improved technology in the future, it is unlikely that we will ever gain complete knowledge of the underworld (or indeed know that we have – proving a negative is always difficult), and thus the problem of ensuring that maximum benefit is gained from existing knowledge is crucial: the estimated annual total of indirect and direct costs of maintaining the nation's underground infrastructure is in excess of £3B [4] – thus even a small improvement could have great benefits.

Knowledge and data integration has long been recognised as an important practical problem, having interesting theoretical aspects too (e.g. [14, 13, 11]). Lenzerini [14] has identified some of the main problems in data integration: 1. Heterogeneity of sources (intensional and extensional level); 2. Limitations in the mechanisms for accessing the sources; 3. Materialized vs virtual integration; 4. Data extraction, cleaning and reconciliation; 5. How to process updates expressed on the global schema, and updates expressed on the sources; 6. The querying problem: How to answer queries expressed on the global schema; 7. The modelling problem: How to model the global schema, the sources, and the relationships between the two. Of particular interest to us here, is the integration of spatially referenced data. There are several reasons why the spatial integration problem is special, including that not all data is necessarily symbolic, and that spatial proximity influences the dependence of one object on another. We believe that constructing an *ontology* of the domain will be key to a successful integration. This term is used by philosophers to describe that branch of metaphysics that deals with the question of 'What things exist?', In recent years 'ontology' has become a buzz word in information science, where it refers to a rigorous formal specification of a vocabulary of concepts and relations. Such specifications are playing an increasingly important rôle in ensuring the integrity of data both within single, information-rich applications and in the transfer of data between applications. There have been a number of proposals for architectures for spatial data integration, e.g. [10, 1] and for specific techniques to integrate, or *conflate* spatial data, whether statistical, e.g. [18], fuzzy/rough-set, e.g. [2] or ontological, e.g. [10]; these techniques specifically attempt to address the issue that in practice data is imperfect, incomplete and noisy and this needs to be considered explicitly, e.g. [20]. Also relevant is research on integrating thematic information, such as [19].

## 3. Inferring Spatio-Temporal Behaviours

As part of a project in the area of *Cognitive Vision*, we have been engaged in building systems that can take low level, noisy, perceptual data, and autonomously build high level descriptions of the behaviours present in video sequences. The domain we have been considering is in "table top space" [16], learning simple table top card games, but in principle the techniques we have developed might be applied in other spaces, including geographic and environmental space. We cluster the low level perceptual data; each cluster then becomes a symbol which is used as part of a relational description of observed behaviour; and inductive logic programming system is then used to induce high level rules which generalise the observed behaviours. Further details can be found in [15, 17].

## References

[1] D. J. Abel, Beng Chin OoI, Kian-Lee Tan, and Soon Huat Tan. Towards integrated geographical information processing. *Int. J. Geographical Information Science*, 12(4):353–371, 1998.

[2] O Ahlqvist, J Keukelaar, and K Oukbir. Rough and fuzzy geographical data integration. *Int J GIS*, 17(3):223–234, 2003.

[3] B. Bennett, Anthony G. Cohn, Paolo Torrini, and Shyamanta M. Hazarika. A foundation for region-based qualitative geometry. In W. Horn, editor, *Proceedings of ECAI-2000*, pages 204–208, 2000.

[4] M Burtwell, E Faragher, D Neville, C Overton, C Rogers, and T Woodward. Locating underground plant and equipment – proposals for a research programme, UKWIR, TR 03/wm/12/4.

[5] A G Cohn, B Bennett, J M Gooday, and N M Gotts. RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, 1:275–316, 1997.

[6] A G Cohn and N M Gotts. A mereological approach to representing spatial vagueness. In J Doyle L C Aiello and S Shapiro, editors, *Principles of Knowledge Representation and Reasoning, Proc. 5th Conference*, pages 230–241. Morgan Kaufmann, 1996.

[7] A. G. Cohn and S. M Hazarika. Continuous transitions in mereotopology. In *Commonsense-2001: 5th Symposium on Logical Formalizations of Commonsense Reasoning*, 2001.

[8] A G Cohn and S M Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1-2):1–29, 2001.

[9] M Egenhofer and R Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991.

[10] Frederico T. Fonseca and Max J. Egenhofer. Ontology-driven geographic information systems. In *ACM-GIS*, pages 14–19, 1999.

[11] Richard Hull. Managing semantic heterogeneity in databases: a theoretical prospective. In *Proc. ACM Symposium on Principles of Databases (Invited Tutorial)*, pages 51–61, 1997.

[12] A. Isli and A.G. Cohn. A new approach to cyclic ordering of 2d orientations using ternary relation algebras. *Artificial Intelligence*, 122(1–2):137–187, 2000.

[13] M Lenzerini. Data Integration: A Theoretical Perspective. In *Proc. PODS 2002)*, 2002.

[14] Maurizio Lenzerini. Data integration is harder than you thought (extended abstract). In *Proc. CoopIS 2001*, volume 2172 of *LNCS*. Springer, 2001.

[15] D. Magee, C. J. Needham, P. Santos, A. G. Cohn, and D. C. Hogg. Autonomous learning for a cognitive agent using continuous models and inductive logic programming from audio-visual input. In *Proceedings of the AAAI Workshop on Anchoring Symbols to Sensor Data*, 2004.

[16] D Montello. Scale and multiple pyschologies of space. In I Campari A Frank, editor, *Spatial Information Theory: a theoretical basis for GIS*, number 716 in Lecture Notes in Computer Science, pages 312–321, Berlin, 1993. Springer Verlag.

[17] P. Santos, D. Magee, A. Cohn, and D. Hogg. Combining multiple answers for learning mathematical structures from visual observation. In *Proc. of the* $16^{th}$ *European Conference on Artificial Intelligence (ECAI-04)*, 2004.

[18] V Walter and D Fritsch. Matching spatial data sets: a statistical approach. *Int. J. Geo. Inf. Sci.*, 13(5):445–473, 1999.

[19] M. F. Worboys and M. Duckham. Integrating spatio-thematic information. In *GIScience*, 2002.

[20] M.F. Worboys and E. Clementini. Integration of imperfect spatial information. *J. Vis. Lang. & Comp.*, 12:61–80, 2001.

# Ontology Alignment for the Semantic Integration of Heterogeneous Geospatial Data Sets

Isabel F. Cruz
Department of Computer Science
University of Illinois at Chicago

## Position Statement

As a growing amount of networked information is becoming available, the potential of a large variety of applications accessing that data in a flexible and transparent way using a single query is far reaching. Particularly important is the case where hundreds of such data sources need to be queried in geospatial, scientific, educational, and e-commerce applications. For example, in homeland security applications, queries on a variety of different domains, e.g., geospatial, biological, and governmental, should be deployed with relatively little effort.

A fundamental issue is the integration of data whose schemas display semantic heterogeneities. We have been working on an approach where no *a priori* integration of all the data schemas involved is needed. Instead, querying can be easily extended to a new data source [3, 5]. Our approach has been used in real-world scenarios in the geospatial domain [1, 4, 6] and leverages current or emerging Semantic Web standards for the storage, interchange, and processing of data.

We have adopted an *ontology-based interoperation* mechanism, whereby a global ontology and the ontologies that describe the distributed data sources need to be *aligned* so as to establish *mappings* between the global ontology and each distributed ontology. A homogeneous view over all the data sources is thus obtained [2, 7]. We have investigated ways to help experts establish such mappings using a combination of manual methods, which are needed for the accuracy of the mappings, and automatic methods, which are needed to facilitate the experts' tasks. Those mappings produce *agreements*, which are declarative and therefore easy to understand and to maintain. Such agreements are automatically incorporated into the queries that result from the translation of the end user's query, which is expressed using concepts in the global ontology.

The *Agreement Maker* is a software tool that we have been designing, implementing, and user testing. It is used to create the mappings between the global ontology and each distributed ontology and to generate an agreement document. Figure 1 shows a user interface that is part of the Agreement Maker. The two hierarchies that represent the ontologies are displayed in two separate panes. The expert browses through the contents of the ontologies and establishes the mappings between each concept (or group of concepts) in the global ontology and the corresponding concept (or concepts) in the distributed ontology. A deduction module is integrated into the Agreement Maker to implement the automatic methods that infer new mappings, which are propagated along the ontotologies [7].

Open research issues include:

- Refinement of the mapping process, by exploring "similarity layers" between concepts and by defining new mapping types that can capture a wider range of semantic heterogeneities.
- Further automation of the mapping process by using other matching techniques, including similarity scores between concepts and multiple dictionaries.
- Experimentation with large ontologies to determine the efficiency of the automation process and its dependency on the topological properties of the graphs that are involved.

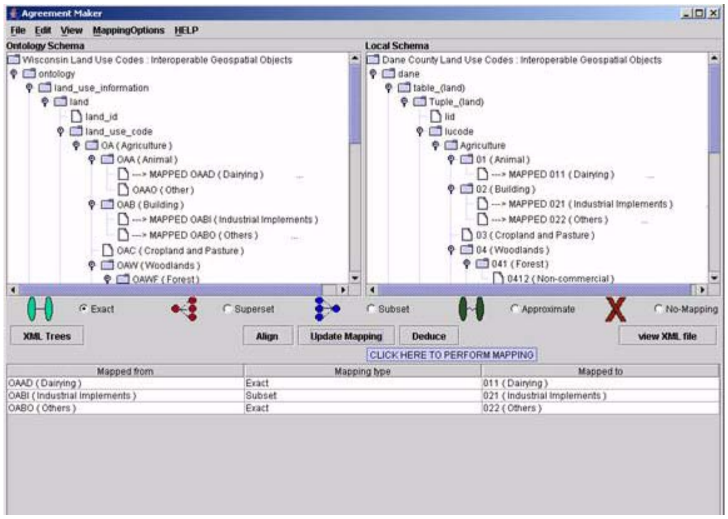


Figure 1: User interface showing established mappings.

- Refinement of the query translation process so as to allow for complex mappings to be incorporated into the queries.
- Extension of the current approach, which deals with a single theme, to data integration across several themes.

## References

[1] I. F. Cruz and P. Calnan. Object interoperability for geospatial applications: A case study. In I. F. Cruz, S. Decker, J. Euzenat, and D. McGuinness, editors, *The Emerging Semantic Web*, pages 281–295. OIS Press, 2002.

[2] I. F. Cruz and A. Rajendran. Exploring a New Approach to the Alignment of Ontologies. In *Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, in cooperation with the International Semantic Web Conference*, 2003.

[3] I. F. Cruz and A. Rajendran. Semantic Integration in Hierarchical Domains. *IEEE Intelligent Systems*, pages 66–73, 2003.

[4] I. F. Cruz, A. Rajendran, and W. Sunna. XML Database Integration for Visualizing US Election Results. In *National Conference on Digital Government Research*, pages 403–406, 2002.

[5] I. F. Cruz, A. Rajendran, W. Sunna, and N. Wiegand. Handling Semantic Heterogeneities Using Declarative Agreements. In *ACM GIS 2002*, pages 168–174, 2002.

[6] I. F. Cruz, W. Sunna, and A. Chaudhry. Ontology Alignment for Real-World Applications (Poster). In *Conference for Digital Government Research*, 2004.

[7] I. F. Cruz, W. Sunna, and A. Chaudhry. Semi-Automatic Ontology Alignment for Geospatial Data Integration. In *GIScience 2004*, LNCS. Springer Verlag, 2004.

# Geospatial Semantic Web Research at LAITS

Liping Di
Laboratory for Advanced Information Technology and Standards (LAITS)
George Mason University
9801 Greenbelt Road, Suite 316-317, Lanham, MD 20706

## Introduction

Geospatial science is the science and art of acquiring, archiving, manipulating, analyzing, communicating, and utilizing spatially explicit data for understanding both physical, biological, and social systems on the Earth's surface or near the surface. Because of their importance in social and economic activities, large amounts of geospatial data have been collected by various public and private organizations mainly using remote sensing methods. Those data must be converted to information and knowledge before they become useful.

## Geospatial Semantic Web

Traditionally, the conversion from geospatial data to knowledge requires the users at their local site having significant amount of domain knowledge on information/knowledge extraction from raw data and on the geospatial data processing. The conversion also requires significant amount of local computer hardware and software resources. As a result, currently the use of geospatial data is still very expensive and most of potential users of geospatial data don't have such luxury. The fundamental problem is that current data and information systems at data providers' site can only provide on-line data ordering and access at best, not the user-specific information and knowledge. We are rich in geospatial data but poor in up-to-date geospatial information and knowledge that are ready to be used by anyone who wants to use. In the past several years, the geospatial community has been worked on solving the problem. The major goal is to make the geospatial information the mainstream information so that anyone can easily obtain the ready-to-use customized geospatial information and knowledge when they want. The major enabling technology is the service-oriented architecture (SOA) and associated interoperable web/Grid service technology.

Geospatial services are the services that handle the geospatial data and information. It is envisioned in the near future, many standard-compliant, interoperable geospatial services will be available on the Web. Those interoperable services form the geospatial semantic web that enables users to obtain their specific geospatial information and knowledge from it. Although there are still many technology issues remaining, the geospatial semantic web is rapidly changing the paradigm of geospatial knowledge discovery from everything owned and operated locally to large-scale sharing of geospatial data, information, software, hardware and other resources over the web.

The Laboratory for Advanced Information Technology and Standards (LAITS), George Mason University, has been working for many years on making the geospatial semantic web a reality. The major work includes leading the development of many federal, national, and international standards in geospatial data and service interoperability and developing advanced information technologies and systems for automating the processes from geospatial data to information and knowledge. This position paper briefs some of LAITS's research projects related to geospatial semantic web.

**Interoperability Standards**

In order for SOA to work, interoperability standards related to all aspects of geospatial service operations at data, information, and knowledge levels are needed. The geospatial community has developed a set of geospatial standards through the standard-setting bodies, including ISO TC 211, the Open GIS Consortium (OGC), and the U.S. Federal Geographic Data Committee (FGDC). Among those standard bodies, OGC is the one primarily concerned with the establishment of implementation specifications for geospatial interoperability. LAITS has been involved in standard-setting activities of all those organizations. It has led or is leading the development of several major standards, including ISO 19130 Sensor and Data Models for Imagery and Grid, ISO Rule for Encoding Imagery and Gridded Data, ISO Radiometric Calibration and Validation of Remote Sensing Data, FGDC Content Standard for Remote Sensing Swath data, FGDC Content Standard for Digital Geospatial Metadata, Extensions for Remote Sensing Metadata, OGC Web Coverage Service (WCS) Implementation Specification, OGC Web Image Classification Service (WICS) IPR, etc.

**Geospatial Web Services**

Recently, LAITS received a grant from NASA REASoN program to conduct research on geospatial web services. The project title is "NASA EOS Higher-Education Alliance: Mobilization of NASA EOS Data and Information Through Web Services and Knowledge Management Technologies for Higher-Education Teaching and Research." This project is developing geospatial web service technologies based on Geo-object and Geo-tree concepts, OGC, ISO, W3C standards for interoperability, and service chaining for construction of complex geospatial models. Through the technologies developed by the project, we will make NASA EOSDIS data and computational resources more easily accessible and usable by higher-education community. A geospatial web service system, called GeoBrain, is being developed by the project. GeoBrain is a distributed, interoperable, federal-able, web-service based geospatial knowledge system. The system will enable anyone with an Internet connected PC to explore the peta-bytes of geospatial data at various data repositories, the huge computing power and scientific algorithms available at NASA and other federal agencies, and the fast network just like they posses such resources. The major components of the system are the geospatial service modules and models that can be integrated just in time for fulfill users' request on geospatial information.

**Geospatial Grid Services**

Grids are persistent environments that enable software applications to integrate instruments, displays, computational and information resources that are managed by diverse organizations in widespread locations. The Globus toolkit is the de-facto standards for Grids. The implementation of SOA in the web environment is called Web services and in the Grid environment the open Grid Services. Currently the web service and grid service are converged with the introduction of Web Service Resource Framework (WSRF). Geospatial Grids are the extensions and domain-specific applications of the fundamental Grid technology in the geospatial discipline. LAITS has been worked on development of geospatial Grids for many years. Currently, LAITS is working on a project for integration of OGC geospatial Web service technology with Grid technology for geospatial modeling and applications. The objectives of the project are 1) to enable the management of geospatial data by Grids; 2) to provide OGC standard compliant access to Grid-managed geospatial data; and 3) enable geospatial modeling and the production of virtual geospatial products in the Grid environment.

**Intelligent Geospatial Knowledge Systems**

The geospatial data, standards, interoperable services, and computational resources form the infrastructure of geospatial semantic web. How to automatically convert geospatial data to user-specific geospatial information and knowledge by using such an infrastructure is one of major research issues in geospatial semantic web. An intelligent geospatial knowledge system should be able to answer many "what if" questions by automatically and intelligently chaining individual service modules to form a complex geospatial model, matching the input data with the model, and executing the model to deliver the answer to the users. Currently, LAITS has three research projects funded by NGA and NASA for studying all aspects of web-service-based intelligent geospatial knowledge systems, from knowledge mining to system architecture and prototyping. The key research areas of those projects are 1) standard-based automated geospatial data and services discovery and access; 2) domain knowledge-driven intelligent decomposition of user query into geospatial processing model for workflow construction; 3) automated geospatial web service chaining, binding, and execution based on the workflow; 4) management of workflows and geospatial models.

# Earth System Science and Spatial Webs (DRAFT 2004-11-20)

James Frew

*Donald Bren School of Environmental Science and Management*
*University of California*
*Santa Barbara, CA 93106-5131*

## Earth system science information

Earth science is increasingly multidisciplinary (i.e., Earth *system* science) and increasingly data-driven:

- Research questions such as elaborating the global carbon cycle require close cooperation between (at least) biologists, geologists, oceanographers, meteorologists, and information scientists.
- Key tools in addressing such questions are the spaceborne remote-sensing systems deployed during the last decade, which are now yielding terabytes per day of raw observations, data rates that were formerly the exclusive domain of supercomputer-hosted models.

Together these transformations are driving the *decentralization* of Earth science information systems:

- As the number of disciplines involved in a research problem increases, the likelihood of the appropriate investigators being co-located decreases.
- As the volume and complexity of data streams increase, the likelihood their being accommodated by a single system decreases.
- There is also the decreasing likelihood that a critical mass of investigators will co-located with a critical mass of data streams (e.g. satellite ground data systems).

"Investigator-led processing", as exemplified by the Federation of Earth Science Information Partners (www.esipfed.org), is the emerging paradigm for both creating and publishing Earth science data and information. That is, the scientists who develop the techniques and algorithms for transforming observations into measurements and measurements into models, are also responsible for generating and disseminating the resulting data and information products. In this new paradigm, Earth scientists are both consumers and producers of data products, assuming roles once reserved for large centralized data centers. Moreover, these distributed science data consumer-producers must be able to *federate* into (possibly *ad hoc*) product chains, where one scientist's product (e.g., snow cover maps) becomes another scientist's input (e.g., to a runoff forecast model).

## Web fundamentals

What does the Web mean for this kind of distributed, federated Earth system science? Recall that, at its most fundamental level, the Web is:

- An object transfer protocol (HTTP), as implemented by a set of Internet servers;
  - i.e., *ubiquitous service*
    - i.e., a *distribution* mechanism
- An object naming scheme (URI), as supported by a set of HTTP servers.
  - i.e., *universal names*
    - i.e., a *federating* mechanism

Therefore, if:

- Earth science information providers individually obtain and disseminate their products through HTTP services;

- Earth science information providers collectively agree on how their products and services are to be referenced via URIs

then we have the simplest possible mapping of Earth system science onto the Web.

Unfortunately even this minimal level of interoperability is nowhere near ubiquitous. There is as yet no universal agreement on:

- formats in which Earth science objects can be encoded for transmission;
- services by which common transformations of these objects can be requested;
- nomenclature for referring to products and services

Formats (e.g., HDF) and services (e.g., DAP, WxS) are receiving by far the most current attention from the computing, Earth, and GI science communities, although since a distributed system must always be prepared to accommodate a multiplicity of either, "victory" in the standards process is less important here.

Naming, on the other hand, is the federating glue that holds a distributed system together. You can't retrieve an object or invoke a service if you don't know what it's called.

The "non-spatial" web has evolved three mechanisms to deal with name discovery:

- search engines (e.g., Google), which look for occurrences of words or phrases in an object's content;
- directories (e.g., dmoz), which categorize URIs according to some particular cataloging scheme;
- informal patterns (e.g., www.*companyname*.com, /~*username*, etc.), which make it easier to guess an unknown URI.

Search engines are of course the most wildly successful of these strategies, but are conspicuously unsuitable for discovering spatial information:

- Most of the search qualifiers for spatial information are non-textual;
- Content-based indexing (and therefore searching) of spatial information is still largely a research problem;
- The most successful search engine (Google) bases its ranking on a particular kind of human-created metadata (the hypertext link) that (currently) has no analogue for spatial data.

Therefore I assume that (for the near-to-medium term, at least), directories and naming schemes will be the primary structures from which a spatial web is built.

## Spatial webs and naming

By "spatial web" I mean "web of spatial information", not "spatialized web of information". The latter leads logically to a "Digital Earth" model: point at a model of the world and access the information associated with the indicated location. While the Digital Earth is compelling, I believe that universal *access* to spatial information is a more immediate (and by no means solved) impediment to distributed federated Earth science. A web of spatial information is a necessary precondition for a successful Digital Earth.

The critical questions facing a "spatial web" are therefore:

- How do we name objects in a spatial web?
- How do we discover these names?

In the interests of space I will limit the following to a discussion of the first question.

The traditional Earth science approach to systematic object naming is to develop project-specific nomenclatures based on unique properties of data or its processing, or both. These kinds of names are called "semantic identifiers" since they encode object-specific metadata. For example MODIS granules managed by the EOSDIS core system are given names that encode the granule's type, version, datetime of acquisition, and datetime of processing. In effect, such names are equivalent to a "title" in a nonspatial cataloging system.

Semantic identifiers have the desirable properties of:

- <>uniqueness within a specific domain
    - <>i.e., the comprise a namespace);
- ease of distributed generation

Withing the MODIS community, the semantics of a MODIS granule identifier reveal crucial information about the granule, and also happen to guarantee uniqueness within the community. Outside the MODIS community, MODIS granule identfiers are opaque. They are not globally unique, but they can be made so if extend with information that identifies that community.

Most of the work on distributed object naming has focused on persistence (e.g. DOI, DRI, PURL), and therefore assumes that semantic identifiers are evil, since the contexts in which the semantics are understood cannot necessarily be preserved along with an object. However, as [Kunze] points out, "persistence is purely a matter of service": all persistent identifier schemes require the persistence of a supporting service infrastructure.

I therefore suggest that the Earth science community federate around a loose hierarchy of naming schemes. The broader community would assign nodes in the hierarchy, leaving the node owners to assign the underlying identifiers as they see fit. For example, if the broader community agrees to use MODIS as a node (or, more specifically, a namespace identifier), then a partial URI of the form MODIS/granuleID becomes a universal granule name. Appended to an appropriate server URL, it becomes a means of accessing services defined for the object (for example, http://server/MODIS/granuleID). The is the idea behind MODster [Frew], in which the sole service provided is redirection.

Reliable naming is especially critical given that many of the objects in a spatial web may be huge (e.g. a single MODIS level 1 granule is several hundred megabytes. [Gray] notes that "Over the last 40 years telecom prices have fallen much more slowly than any other communication technology", making it ever cheaper to move processing to data rather than the other way round. Eventually all spatial data may well be exposed through high-level services that implement any reasonable custom processing, but until then, it will be very important locate the most accessible copy of a large object, or especially the copy that happens to reside at a site that also implements the required level of custom processing. This kind of redirection to an appropriate data/service combination is straightforward with a approach like MODster.

## References

[Frew]
Frew, James and Gallagher, James (2002) MODster: Peer-to-Peer Sharing for Remote Sensing Standard Products. IGARSS '02 International Geoscience and Remote Sensing Symposium. IEEE Computer Society, Toronto, Canada.

[Gray]
Gray, Jim (2003) Distributed computing economics. Microsoft Research Technical Report MSR-TR-2003-24.

[Kunze]
Kunze, John (2003) The ARK persistent identifier scheme.

# A situated representation of GIS resources and GIS users

Mark Gahegan
GeoVISTA Center, Penn State Geography

In GIS and most other computational systems, datasets, methods and other electronic resources are treated as objective entities. However, they were created for, and are applied to, specific situations. These situations are generally not captured or represented in any way, except perhaps in the minds of the people involved. And when data producers and data consumers do not share such situational knowledge, the chances of misunderstanding and misapplication increase. Such a situation is very common among agencies and organizations whose primary role is in the creation of information produces, such as NGA.

Some aspects of the situations surrounding creation and use of resources can be harvested, remembered, mined, visualized and applied to help increase subjective understanding, and to compliment the more objective, top-down knowledge that might be provided by computational ontologies. The result is <u>an intricate web of relationships linking people, methods, data, places, times, concepts, tasks and more that builds up over time as resources are created, modified and used by communities of researchers.</u> Such provenance information can be accumulated over time via a coordinated GRID infrastructure that can log specific use-cases to a central node and a harvesting system that uses association rule mining to isolate trends and significant relationships. Resources are thus contextualized in a manner that reflects, to a limited extent, similarities and differences in the way they are understood by their users, *through* the way they are used. Users can choose to commit to various levels of logging, and can also choose with whom they wish to share both the resources they produce and information concerning the resources they employ.

The resulting webs can also contain more formal knowledge, such as concept maps and ontologies, and in fact these resources receive exactly the same treatment as datasets and methods in that they can be tracked and contextualized through their use by a community of scientists or analysts. Procedural or task knowledge can also be added, so that user workflows and logged actions can be included, in which case the entire history of how a product or method was created and how it has since been used can be explored and utilized for a variety of ends: Who made this? How? Who has used it? How? Who has modified it?, etc.. The (spatial) webs are thus an amalgam of resources, formal descriptions, informal situations (use cases), procedural knowledge, social networks of people and a geographical context involving space, time and scale.

Some of these ideas are demonstrated via a Web Portal (Codex) that forms a gateway to e-resources for groups of researchers working together. Specific examples are drawn from the user-communities within the Geosciences Network (GEON: www.geongrid.org) and the Human-Environment Regional Observatories (HERO: http://hero.geog.psu.edu). Examples are used to show how the knowledge derived from situations can be utilized to help contextualize unfamiliar resources.

# Semantics of What?

Werner Kuhn
Institute for Geoinformatics
University of Münster

*Position Paper*
*NCGIA Specialist Meeting on Spatial Webs*
*Santa Barbara, CA, December 2-4, 2004*

Semantic interoperability is a noble and important goal, but hard to pin down, for several reasons: the notion itself is somewhat redundant, its meaning is elusive as is that of its parts (semantics, interoperability), there is no commonly accepted formal definition, there are no benchmarks and no agreed challenges, the role of humans in the loop is unclear, and the acronym inflation around the semantic web obscures rather than highlights many research issues. As a consequence, semantic interoperability research often doesn't know where to start from, where to go, and how to get there. This position paper tries to determine a starting point of the research by asking *"what is it that needs to be semantically defined in order to support semantic interoperability?"* The intention is to encourage more precise problem statements in semantic interoperability research.

## Semantic Interoperability

Before we can determine what it is that needs semantic specifications, we should clarify what we expect to interoperate. In the real world, it is hard to imagine two agents interoperating successfully without a shared understanding of the messages they exchange. If you and a friend prepare a meal without a shared understanding of expressions like "a cup of" or "simmer until tender", no edible food will result. Yet, the kind of interoperability currently supported by standards like those of the Open Geospatial Consortium (OGC) is roughly at this level: interface standards establish syntactic protocols for invoking system behavior, but do not specify in machine-readable form the meaning of the terms used in the protocols. Consequently, OGC tests individual components for conformance to its specifications, but not multiple components for interoperability - whatever that would mean, with or without the "semantic" qualifier. The need to attach semantics to the syntactic protocols has been well recognized in OGC since its early days. Yet, the semantics work of OGC seems to have stalled after establishing its initial grand vision of interoperating information communities and semantic translation.

## Interfaces

The early focus of OGC on service interfaces has been very wise. From a semantics point of view, one wishes that this focus had not been lost in the recent shift of attention to GML-based data exchange. But it has: major OGC efforts are now going into database schema harmonization, and the oxymoron of "data interoperability" has crept into industry and agency jargon[1]. It suggests that the flower and eggs in your kitchen interoperate among themselves to prepare your meal. My point is not that humans need to be involved, but that it takes *operations* to interoperate, not just data.

For over 30 years, software engineers have known that data and operation semantics can only be captured together, through interfaces of software modules (Parnas 1972). The idea of a component interface is an ideal (and I would claim the only sound) basis for understanding semantic interoperability. Agents, computational or human, interoperate through interfaces. For them to interoperate in a meaningful way, these interfaces need well-defined semantics. Thus, semantic interoperability is really the only kind of interoperability.

It is essential to consider the human, organizational, and societal issues involved in information sharing and integration. However, it is not helpful to muddle the definition of (technical) interoperability through service interfaces with these aspects. We need clear, crisp, and measurable interoperability criteria if interoperability is to be a research goal. These criteria should be defined incrementally at multiple levels, starting at the technical and proceeding through human, organizational, and social levels.

How can we define interoperability at a technical level? Within and outside the geospatial community, a definition given by ISO TC204 (document N271) is often quoted:

> *"The ability of systems to provide services to and accept services from other systems and to use the services so exchanged to enable them to operate effectively together."*

While this definition is technical enough to be useful in systems engineering and testing, it falls short of establishing verifiable criteria. What does it mean for systems to operate together? And when can they be said to do this effectively? A more precise and verifiable definition of interoperability requires two things:

1. going back to the basics of what software interfaces are and what they mean
2. giving a mathematical definition of interoperability, based on interface specifications.

This paper addresses the first requirement. The second is pursued by people like Joseph Goguen, Michael Grüninger, Robert Kent, Till Mossakowski and Marco Schorlemmer. Preliminary results of an ongoing debate can be found in these colleagues' discussion papers at http://www.dagstuhl.de/04391/Materials/. One can gather from there that the mathematical theory of institutions (Goguen 2004 (draft)), based on category theory, is expected to supply the necessary formal foundations. It rests on *signatures* and establishes the core semantic notion of models satisfying sentences built from signatures in any language. A key question is whether these models should simply be sets (as in classical model theory) or have some algebraic structure imposed by functions.

This paper also does not talk about that other interface: the one between humans and systems. It is equally, if not more important in the quest for semantic interoperability than the machine-level interfaces. The basic point about interface semantics remains the same, but gets complicated by the difficulty to frame human-computer interaction as a formal language, and by the connection to natural language semantics.

[1] See, for example, http://www.esri.com/software/arcgis/extensions/datainteroperability/

## Signatures

The interface of a software component (such as a web service or a Java package) is formally captured by a *signature*. A signature describes a component's type information, consisting of the input type(s), output type(s) and names of the operations offered by the component. An example of a signature is the specification of distance operations given in the OGC/ISO Spatial Schema standard:

**GM_Object :: distance (geometry : GM_Object) : Distance**

This signature says that the distance operation is applied to a geometric object (GM_Object), takes as input another such object (called geometry), and returns a value of type Distance.

More generally, the term signature refers to a collection of such operation signatures. Signatures are basic elements of algebra and algebraic software specifications (Ehrig and Mahr 1985). Formally, they contain *three kinds of symbols*, standing for

- functions (of a certain arity)
- sorts (types), and
- constants (tokens).

In the distance example, the only function symbol is **distance**, the sort symbols are **GM_Object** and **Distance**, and there is no symbol for constants. A more mathematical form of the example signature would thus be

**distance** : **GM_Object** x **GM_Object** → **Distance**

Note that this form treats the two geometries symmetrically, as it should. A shortcoming of today's software specifications and implementations is that they rarely provide for this decoupling from single object classes.

## Semantics

The only sensible use of the term 'semantics' refers to *expressions in a language*. Such expressions can consist of individual symbols ("words") or symbol combinations. Neither concepts nor entities nor properties nor processes have semantics, but the symbols of the languages describing them do (or need).

In an information system context, such languages are manifold: programming languages, schema languages, query languages, interface specification languages, workflow modeling languages, user interface languages, sensor signals, and others. Many of these languages are extensible, i.e., they allow users to introduce new symbols (for data types, attributes, relationships etc.). Additionally, information system standards introduce all sorts of more or less controlled vocabularies (such as those of feature-attribute catalogues or metadata standards). Furthermore, free-form text entries open the gate to almost unlimited uses of natural language expressions in geospatial information sources.

Coping with geospatial semantics means, eventually, building ontologies specifying the meaning of expressions in most or all of these languages. Semantic interoperability research, however, allows us to focus on a very small subset of languages: those defining and using service interfaces. According to the previous section, interfaces are signatures, containing three kinds of symbols as parts of a well-defined structure. Thus, *the semantics required for semantic interoperability is that of expressions built around service signatures*.

## Semantics of Signatures

Various languages serve to specify service signatures, to express calls to services or results from them and to reason about services. In the semantic web context, such languages are used to describe, discover, evaluate and invoke web services. For example, WSDL (the Web Service Description Language) allows for syntactic descriptions of web service interfaces and OWL-S (the service ontology of the Web Ontology Language) has been proposed for semantic specifications of services. More comprehensive service modeling efforts like WSMF (the Web Service Modeling Framework) are currently under way (Fensel and Bussler 2002).

Regrettably, the ways in which signatures obtain semantics in these languages are far from clear. Since most ontology languages privilege static entities and binary relationships over more general relationships and processes, it is hard or impossible to say meaningful and useful things about the functions performed by services. The algebraic structure imposed on domains by functions is typically lost. The standard semantic web approach is to specify input and output types, pre- and post-conditions on them, and taxonomies of service types. While input and output specifications can simply point to domain ontologies, pre- and post-conditions require richer models of the domains as well as languages to express rules. The necessary expressiveness and formal bases of these languages are not yet understood, the granularity of service type taxonomies is too coarse, and the service types themselves are not semantically defined.

In other words, while semantic web languages can say something about sets of values or objects in a domain, they are either silent or confused about the mappings between these sets established by services. *Service semantics has no sound formal basis in today's semantic web.* We simply don't know yet what needs to be said (and how) about the semantics of signatures to make the corresponding services semantically interoperable. The suggestions in the semantic web community that a few more half-baked W3C standards will solve these problems are rather optimistic. Much more research is needed on the foundations of service semantics. This research needs to be guided by non-trivial application scenarios.

The case of "spatial webs" offers tremendous application challenges, as well as some solid ground in the form of geospatial information standards. But is spatial information special when it comes to interoperability?

## Geospatial Semantics

In the absence of a general theory of service semantics, it is hard to state clearly why and how geospatial services would be special. At the level of establishing semantics for service signatures, there seems not to be anything special about space (and time). Yet, geospatial information and its semantics are characterized by some properties that could guide our search for ways to achieve semantic interoperability. I will close with some guesses about these properties:

1. Geospatial data and services contain symbols whose meaning is not only a matter of convention. For example, a wind direction returned by a weather service, or a water level measured by a gauge have an *observable grounding in the physical world*. Conversely, the meaning of their measuring units, of a currency amount, or of a single-click purchase at Amazon is purely conventional.

2. Because of this physical grounding, explaining the semantics of geospatial information will require elaborate *measurement ontologies*. The emerging sensor web technology will only be useful, if such ontologies are widely available and carefully tied to existing standards in science and engineering.
3. At the same time, geospatial information is often based on *human perception and social agreements*, rather than objective measurements. Coping with the meaning of qualitative judgments (say, of landscape aesthetics) or of social constructions (like neighborhood classifications), and providing mappings among such categories, without imposing incompatible abstractions, are probably the biggest challenges to make geospatial information more meaningful and shareable.
4. A special case of social agreements are *geographic names* and other *identifiers* of geospatial entities. Geographic name registries in the form of gazetteers will need better translation and geo-referencing capabilities. Object identifiers in different databases across information communities will need to be linked. For example, the same petrologic sample could be registered under different identifiers and reference to different geographical names in various online databases supporting geochemical analyses.
5. Space and time are primarily understood through *processes*: we locate stuff because we can move it (not the other way round!), we use distances and directions to navigate, and we determine when to leave the beach by the estimated speed of the advancing storm. This process-nature of geospatial information challenges the entity-bias of the semantic web.
6. Last but not least, *vagueness and different levels of granularity* are fundamental to geospatial information. Theories of vagueness and mappings among granularity levels of geospatial ontologies are therefore essential ingredients of spatial webs.

One approach to these and other requirements is to generalize the notion of spatial reference systems toward *semantic reference systems* (Kuhn 2003). Their task would be to help interpret and translate geospatial information in general, rather than just coordinates. We have only begun to study the computational (Kuhn and Raubal 2003) and institutional challenges posed by this vision. However, pursuing this idea seems a logical consequence of the observation that space itself acts as an integrator of information, through both, locations and the phenomena observed at them.

## Conclusions

What is it that needs to be semantically defined in order to support semantic interoperability? This paper has claimed that it is service signatures, which are established notions in mathematics and software engineering. However, it found a lack of understanding how to specify the semantics of signatures in the context of web service description and discovery. And it observed an abandoning, in practice, of service interfaces as the core elements of interoperability, in favor of coarse-grained generic interfaces (like those of map or feature servers) and database schema exchange.

Is a call for fine-grained theories of semantics, at the level of operation signatures, still justified? It may well be that the frustration with fine-grained *syntactic* specifications coming out of the CORBA world was a sign of the need for coarser grained alternatives in semantics as well. On the other hand, we should not throw out the semantic baby with the syntactic bathwater. As long as nobody proposes a mathematical foundation for coarser grained semantic interoperability, we either stick to the small grain, which has such a foundation, or give up on semantic interoperability as a scientific endeavor.

## References

Ehrig, H. and B. Mahr (1985). *Fundamentals of Algebraic Specification*. Berlin, Springer-Verlag.

Fensel, D. and C. Bussler (2002). "The Web Service Modeling Framework WSMF." *Electronic Commerce: Research and Applications*(1): 113-137.

Goguen, J. (2004 (draft)). Information Integration in Institutions. (Memorial volume for Jon Barwise). L. Moss. Draft available from http://www.cs.ucsd.edu/~goguen/papers/IF.html.

Kuhn, W. (2003). "Semantic Reference Systems." *International Journal of Geographic Information Science (Guest Editorial)* **17**(5): 405-409.

Kuhn, W. and M. Raubal (2003). Implementing Semantic Reference Systems. AGILE 2003 - 6th AGILE Conference on Geographic Information Science, Lyon, France, Presses Polytechniques et Universitaires Romandes.

Parnas, D. L. (1972). "A Technique for Software Module Specification with Examples." Communications of the ACM **15**(5): 1053-1058.

# Overcoming Differences of Meaning during the Discovery and Retrieval of Geospatial Information in Spatial Data Infrastructures

Michael Lutz, Eva Klien

Institute for Geoinformatics, University of Münster
Robert-Koch-Str. 26-28, 48149 Münster, Germany

Advances in sensor technology are revolutionizing the way that geospatial information is collected and analyzed. Already today (and even more so in the foreseeable future) sensors provide continuous streams of geospatial information (GI). While this opens opportunities for improving the way decisions are taken in a variety of fields, it also presents a number of challenges. We are focusing on the crucial question of *how to find suitable information* for a given task (GI discovery) and *how to access it* once it has been found (GI retrieval).

Spatial Data Infrastructures (SDIs) provide searchable catalogs of information descriptions (metadata) in a standardized format (ISO/TC-211 2003) and through a standardized interface (OGC 2004). For GI retrieval, too, standardized interfaces are provided and widely used in the form of Web Feature Services (OGC 2002). However, a number of problems are caused by differences in meaning (semantic heterogeneity) during both GI discovery and retrieval.

Current GI discovery is largely based on string-matching keywords or other search terms. Even though natural language processing techniques can increase the semantic relevance of search results w.r.t. to the search request (e.g. Richardson & Smeaton 1995), keyword-based techniques are inherently restricted by the ambiguities of natural language. As a result, keyword-based search can have low recall if different terminology is used and/or low precision if terms are homonymous or because of their limited possibilities to express complex queries (Bernstein & Klein 2002).

Once a suitable information source has been discovered and could be accessed through a WFS interface, ambiguity still presents a problem. While it is possible to obtain syntactic descriptions of the application schema of the feature types, this description is often not sufficient for interpreting the meaning of its attributes. This makes it difficult (or in some cases impossible) to create a query expression that actually retrieves the desired information.

To overcome these problems, we propose to use ontological (i.e. explicit, formal and widely-agreed) descriptions of information sources. Thus, the semantics of the content of information sources become machine-interpretable, and users are enabled to pose concise and expressive queries. Furthermore, logical reasoning can be used to discover implicit relationships between search terms and information descriptions as well as to flexibly construct taxonomies for classifying information sources (Klien *et al.* 2004).

In recent years, a number of similar approaches have been proposed (e.g. Lin *et al.* 2003). However, they have not been widely adopted outside academia yet. We believe that a number of challenges need to be addressed before ontological approaches will reach widespread adoption.

- *Integration into existing SDI architecture.* To facilitate adoption, it should be possible to easily integrate ontology-based approaches for GI discovery and retrieval into existing SDI architectures. This is preferably realized by extending existing components and adding new components where necessary. For example, in (Klien *et al.* 2004) we present a component that understands queries that contain ontological query concepts. It accesses a reasoning component to enrich the query with additional ontology concepts and then sends the enriched query to a conventional catalog service. As the "enhanced" catalog service reuses the standardized catalog service interface it can also deal with normal (non-semantic) queries.
- *Standardization of domain vocabularies.* Ontology-based approaches overcome some of the problems of conventional metadata standards by enabling expressive but flexible descriptions. However, standardization is still necessary at the level of the used domain vocabulary in order to enable

interoperability *between ontological descriptions.* In the geospatial domain, a body of rich knowledge models already exists in the form of ISO TC 211 and OGC standards, which could be used as the basis for developing an already widely agreed domain vocabulary. Unfortunately, this knowledge is only semi-formal (as text or UML models). How to extract and represent this knowledge in a formal way is therefore an important research issue (Probst *et al.* 2004).

- *User support for creating semantic descriptions of geographic information.* The ontology-based approach of GI discovery and retrieval is only feasible, if all information resources registered in the SDI are semantically described. Taking into account how difficult it is to get data and service providers to specify *conventional* metadata, user support for generating *semantic* descriptions is definitely a crucial issue. This includes methods for automatically extracting (parts of) these semantic descriptions as well as intuitive user interfaces that hide the complexity of ontology languages from the provider.
- *Hiding ontologies and reasoning from the user.* Similarly, SDI users need to be supported in formulating their queries. As understanding and using ontology languages is beyond normal users, an intuitive query language and/or graphical user interface should allow a requester to intuitively formulate a query using a well-known domain vocabulary. As a first step in this direction, we are currently developing a prototype, which provides the requester with a GIS-like query interface as well as a SQL-like query language.
- *Support (semi-)automatic composition of complex processing services.* Finally, if Spatial Webs are also to include processing capabilities, it should be possible to seamlessly combine several data sources and processing services in order to provide a user with an answer to his actual question. Ideally, the composition should go unnoticed by the user. This vision would require methods for semantically describing services and for using these descriptions for service discovery (Lutz 2004). Also, it should be possible to automatically generate mediators based on the semantic descriptions that bridge heterogeneities between the services within such a composite service.

## References

Bernstein, A. & Klein, M. (2002): *Towards High-Precision Service Retrieval*, in: Horrocks, I. & J. Hendler (ed.): The Semantic Web - First International Semantic Web Conference (ISWC 2002): 84-101.

ISO/TC-211 (2003): *ISO 19115:2003. Geogaphic information - Metadata*, International Organization for Standardization.

Klien, E., Einspanier, U., Lutz, M. & Hübner, S. (2004): *An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information*, in: Toppen, F. & P. Prastacos (ed.): 7th Conference on Geographic Information Science (AGILE 2004): 179-188.

Lin, K., Ludäscher, B., Brodaric, B., Seber, D., Baru, C. & Sinha, K. A. (2003): *Semantic Mediation Services in Geologic Data Integration: A Case Study from the GEON Grid*, Geological Society of America (GSA) Annual Meeting 35 (6).

Lutz, M. (2004): *Non-taxonomic Relations in Semantic Service Discovery and Composition*, 1st "Ontology in Action" Workshop, in conjunction with 16th Conference on Software Engineering and Knowledge Engineering (SEKE 2004): 482-485.

OGC (2002): *Web Feature Service Implementation Specification, Version 1.0.0 (OGC Implementation Specification 02-058)*, Open Geospatial Consortium.

OGC (2004): *Catalogue Services Specification, Version 2.0 (OGC Implementation Specification 04-021r2)*, Open Geospatial Consortium.

Probst, F., Gibotti, F., Pazos, A., Esbri, M. A., Benigno, M., Gutiérrez, M. & Kuhn, W. (2004): *Connecting ISO and OGC Models to the Semantic Web*, 3rd International Conference on Geographic Information Science.

Richardson, R. & Smeaton, A. F. (1995): *Using WordNet in a Knowledge-based Approach to Information Retrieval (Technical Report CA-0395)*, Dublin City University.

# Semantics of Landscape: Providing Information About the Landscape via The Semantic Web

David M. Mark
NCGIA & Geography
University at Buffalo, NY 14261 USA

Semantics is all about meaning, and semantic interoperability is about the preservation of meaning as information is exchanged among software, databases, and people. Words in natural language play a critical role in the transmission of meaning from one person to another, in the sharing of concepts across individuals within speech communities. Recent developments in ontologies for knowledge representation also have words and their meanings playing a central role, and mainly formalize the definitions of and relations among words (terms) for a specific domain. The key role of words and language in semantics is recognized in the ubiquitous semiotic triangle, which relates signs (symbols, including words), concepts (mental ideas), and referents (in the real world). The Semantic Web promises to take the world-wide web to new levels of utility by building upon ontologies, inference procedures, and definitions, but if universally agreed-upon definitions are not available, 'Tower of Babel' problems will still occur.

The natural landscape is a very important domain for human activities, and is particularly important in geospatial intelligence. The landscape may be loosely defined as the larger elements of the human environment, those elements that can be perceived and understood from a distance of hundreds of meters or more, such as hills and valleys, lakes and rivers, forests and deserts. However, representation and exchange of information about the landscape in word-centered ontologies such as those alluded to above, for delivery through the Semantic Web, presents a number of significant research challenges that are rare or absent for other domains.

- **Semantics usually deals with objects but geospatial information is often fields or images**. While the semantic web is word-centered, the great majority of existing geospatial information about natural landscapes is in the form of fields (digital elevation models) or images (mainly from remote sensing). Serving of field-based or image-based geospatial information via the semantic web will require robust computational methods for identifying and delimiting landscape features present implicitly in those fields and images.
- **Classifications of landscape elements vary across cultures and languages**. For inorganic natural domains such as landforms and water bodies, natural variation usually forms a continuum, and categories for landscape elements thus have a somewhat arbitrary component, contingent on landscape variation, lifestyle of the people, and linguistic and cultural history.
- **Dependency of landscape objects on classifications**. For many domains, the entities dealt with are bona fide objects that exist as objects independently of their definitions. For example, individual higher organisms such as birds or butterflies are easily perceived as 'things' even if not identified as to type. In contrast, many landscape elements are not bona fide objects, but are parts of a continuous land

surface, and thus different definitions of landscape element types may lead to a different subdivision of the land into objects.

- **The above issues interact!** The same landscape, represented by the same field-like data models, may produce quite different delimited objects and object classes according to the definitions of words in different languages. Feature-extraction methods probably will need to be language-specific. A word-based ontology of a type of natural landscape developed on the basis of one natural language will likely be inadequate for other languages.

The extent of these 'problems' is not yet known but several solid examples suggest that these issues will be a substantial impediments to the development of multilingual semantic interoperability for landscape data using the word-centered approaches now common for knowledge-representation ontologies and for the Semantic Web.

# Proposal to participate in the NCGIA specialist meeting on Spatial Webs

Martin Raubal

Institute for Geoinformatics, University of Muenster
Robert-Koch-Str. 26-28, 48149 Muenster, Germany

## *Cognitive Semantic Interoperability for Spatial (Web) Services*

Achieving semantic interoperability for geographic information has been a major research goal of the GIScience community. It is defined as the capacity of (geographic) information systems and services to work together without the need for human intervention (Harvey *et al.* 1999). Nevertheless, these systems and services eventually support human users in their decision-making and therefore people's understanding of terms must be accounted for (Miller forthcoming). Current approaches to solve the semantic interoperability problem are based on realist semantics, which defines meaning independent of a human user. In order to account for human concepts a cognitive semantics approach is needed. In this position paper we make a case for *Cognitive Semantic Interoperability*.

Let's assume you are using a wayfinding service that communicates wayfinding instructions that are enriched by landmarks. The service automatically extracts salient features from datasets based on a model of landmark saliency (Raubal and Winter 2002; Nothegger *et al.* 2004). By extracting such features it is assumed that the system's concept of a landmark equals the user's concept of landmark. But this is most often not true. It is therefore necessary to represent and to translate between these different conceptualizations. Here we argue for a formal cognitive semantics approach, which allows for capturing the differences between a system's view and a user's view of the same concept. In the long run this will lead to cognitive semantic interoperability between geospatial (web) services and their users.

Realist semantics asserts that the meanings of expressions are in the world and therefore independent of how individual people understand them. In cognitive semantics, meanings are claimed to be mental entities—mappings from expressions to conceptual structures—which themselves refer to the real world. It has been demonstrated that realist approaches to semantics face a number of difficulties, most notably their problems to deal with learning, with mentally constructed objects that have no direct correspondence in the real world, and with the fact that the meaning of concepts often changes both over time and in different contexts (Gärdenfors 2000). During the hype of creating automated information services—services that work with each other without human intervention—it is often forgotten that the final goal is to create information for a human user. Therefore it is on the one hand important how individual people understand the output of a computer system and on the other hand that the computer system understands people's meanings of terms used to formulate a question to the system.

Our approach to allow for cognitive semantic interoperability utilizes Gärdenfors' idea of *conceptual spaces*—sets of quality dimensions within a geometrical structure (Gärdenfors 2000). For example, a color space may be represented by its quality dimensions hue, saturation, and brightness. In (Raubal forthcoming 2004) we describe a methodology to formalize conceptual spaces as vector spaces. A concept is represented as an n-dimensional conceptual vector space, whose axes represent quality dimensions or further vector spaces denoting domains. In order to standardize the variables represented by these axes a statistical method called z-transformation is applied. This allows for representing instances of a concept as points in the conceptual vector space and measuring the distances between them and their distances to a prototype of the concept. Such prototype is ideally an n-dimensional region in the conceptual space. In addition, one can account for the use of a concept in different contexts by assigning different weights to the axes of the conceptual vector space. Such spaces can be utilized for knowledge representation and sharing, and support the paradigm that concepts are dynamical systems. In order to bring the system's semantics closer to the user's semantics, conceptual vector spaces can be mapped from one to another in the way of transformations and projections.

How does this help in the wayfinding scenario? After defining quality dimensions for the concept landmark from both the system's (i.e., the system designer's) and the user's perspectives, the conceptual vector spaces can be formally represented. For example, some wayfinders prefer buildings as landmarks and the quality dimensions may be size, shape, color, and visibility. During wayfinding at night the weight for color can be set low, etc. Both spaces can then be compared—e.g., by representing one's prototype within the other space and determining the distance between the prototypes—and adjusted to each other. For practical reasons, the system adapts the semantics of its concepts to the user's semantics, leading to improved human-computer interaction.

Following a cognitive path to overcome the differences of meaning will be a major interdisciplinary endeavor, involving both spatial and cognitive scientists. With the presented approach many questions are to be solved in the near future, such as:

- How to find the covariances between dimensions and account for these in the representations of conceptual spaces?
- How to formally define a comprehensive formal framework for mappings between conceptual vector spaces and to quantify the loss of information?
- How to pragmatically implement the approach in the context of geospatial web services?

Computer friendly representations in the area of cognitive semantics are exceptional but urgently needed. They allow us to account for the fact that different people have different conceptualizations of the world and therefore require different presentations of answers to their spatio-temporal queries. The concepts of system designers have to be matched with the concepts of individual users to make information communication a successful enterprise. This will lead to useful and usable spatial semantic webs.

## References

P. Gärdenfors (2000) *Conceptual Spaces - The Geometry of Thought*. Bradford Books, MIT Press, Cambridge, MA.

F. Harvey, W. Kuhn, H. Pundt, Y. Bishr, and C. Riedemann (1999) Semantic Interoperability: A Central Issue for Sharing Geographic Information. *Annals of Regional Science* 33 (2)(Geo-spatial data sharing and standardization): 213-232.

H. Miller (forthcoming) What about people in geographic information science? in: D. Unwin (Ed.), *Re-Presenting Geographical Information Systems*. Wiley.

C. Nothegger, S. Winter, and M. Raubal (2004) Selection of Salient Features for Route Directions. *Spatial Cognition and Computation* 4(2): 113-136.

M. Raubal (forthcoming 2004) Formalizing Conceptual Spaces. in: *Formal Ontology in Information Systems, FOIS 2004*, Torino, Italy.

M. Raubal and S. Winter (2002) Enriching Wayfinding Instructions with Local Landmarks. in: M. Egenhofer and D. Mark (Eds.), *Geographic Information Science - Second International Conference, GIScience 2002, Boulder, CO, USA, September 2002. Lecture Notes in Computer Science* 2478, pp. 243-259, Springer, Berlin.

# Semantics of Spatial verbs and their role in interoperability

**Sumit Sen**

Institute for Geoinformatics, Robert Koch Str. 26-28,
48149 Münster, Germany

## The meaning of spatial verbs

Perhaps verbs occur in English sentences as the most important critical component of sentence meaning and completeness. It denotes the notion of process, action or activity and also has a temporal component (tense). There are also verbs which have some intrinsic spatial meaning. These include motion verbs and existence verbs[1] such as {move, fly, drag, pull} and {appear, locate, remain}. For reasons explained in the footnotes we chose to focus on the former category of verbs in our venture of studying spatial verbs.

So what is special about spatial verbs? And are there purely "spatial" verbs? The answer is perhaps no. Spatial verbs are spatio temporal verbs and do not exist in isolation. The tense notion and activity/action distinction exists in a every textual description of spatial behavior. "Cars move on Roads", "Rivers flow into lakes". We take into consideration and still for simplicity sake, use only the spatial sense of these verbs for discussion. We also use *to move* as our example since it occurs as a semantic prime as in the current model of the NSM [1] Further if we use the physical sense of move (as it exists in classical physics) we know of the types of moves based on state-of-rest, steady-motion state and accelerating (or decelerating) state[2]. This allows us to relate the verb corresponding to motion definition in reference to spatial entity types in different dimension cases.

The next step is to use spatial verbs that describe these states. It can be considered inappropriate to use diagrams for representing actions and activities and so it is necessary to use alternate representations such as formulae and physical quantities (like displacement, velocity and acceleration) to represent each case and provide verb forms [3]that have meanings which correspond. Such grouping allows highlighting relative semantic similarity of geospatial behavior concepts and also the differences. For example consider the example of *walk*, *swim* and *run*. All though of them relate to the *move* verb, the run has the additional qualification of higher velocity (word-net has in fact links run to the move concept in two steps as compared to *swim* and *walk*)

The next level of categorization would involve the types of spatial entities involved. Thus for a point feature to *move* on a surface feature could correspond *to swimming* while a point *moving* in a container feature can constitute *to sink.* What is essentially important to note is that there are image schematic differences in the way the verbs are described. It is almost clear in most cases which feature types cannot be involved in a certain type of verb.

## "Decomposable units" and "basic spatial verbs"

[1] The "to be" verb can also be argued as an existence verb but is used in great variety of ways including "is a" or "is a part of" relations. While some of these can be spatial it becomes difficult o dissociate metaphorical extensions of the spatial case.
[2] One can also argue for rate of change of acceleration to account for further states like steady acceleration, Simple harmonic motion and others. Again for simplicity we assume steady acceleration.
[3] WordNet lexicology is used to reference the sense of the verb used.

One natural question would be whether there can be an exhaustive list of basic spatial verbs. Even if the answer is positive it does not provide the purpose of getting atomic concepts of spatial behavior because language is culture and domain dependent. Perhaps the answer lies in cognitive linguistics and embodied lexical development. The work on executing schemas [2] demonstrates an example of what can be constituted behavioral elements and usually expressed as Petri-net diagrams. Our position in regard to composition of behavior using spatial verbs is that the basic verbs should have unique image schematic references. This is not a simple problem because image schemata are extendable. However it is our view that currently available theory is able to provide the basis for most common verbs

The sequels to such questions are related to the existence of semantic primes. Our argument towards the existence of decomposable units of basic verbs is formed on the basis of Jackendoff's proposition on how meanings can be decomposed [3]. It is therefore assumed that there can be many meanings of the same verb when combined with other words (which can be verbs themselves)

**Semantic Interoperability**

Finally what does this all mean in the context of semantic interoperability? Aren't spatial behaviors expressed sufficiently in ontologies? Is there any major assistance or advancement that this approach offers to the interoperability problems like naming or cognitive heterogeneity?

Maybe there are. This is because of our hypothesis that behaviors of objects actually define the concept of the object rather than their attributes. Concepts expressed in most geospatial ontologies used for semantic interoperability do not fully specify concepts with knowledge about its behavior. This inhibits expression of behavior of geographic feature types which an important part of geospatial domain knowledge and is an important link that can help in translation [4] of concepts or even establishing the fact that a translation is not possible.

**References:**

[1] Goddard, Cliff. (2002). The *search for the shared semantic core of all languages*. In Cliff Goddard and Anna Wierzbicka (eds). Meaning and Universal Grammar - Theory and Empirical Findings. Volume I. Amsterdam: John Benjamins. pp. 5-40.

[2] Bailey, D. R., Feldman, J. A., Narayanan, S., and Lakoff, G. (1997). *Modeling embodied lexical development*. In Proceedings of the 19th Cognitive Science Society Conference.

[3] Jackendoff, R. (1990) *Semantic Structures*. Cambridge, MA: MIT Press.

[4] Kuhn, W. (2003). *Semantic Reference Systems.* International Journal of Geographical Information Science.

# Spatial Data Mining and Geo-spatial Interoperability

*Shashi Shekhar*
Computer Science Department, University of Minnesota
200 Union Street #4192, Minneapolis, MN 55455.

## EXTENDED ABSTRACT:

Spatial data mining systems (SDMS) [1] extracts previously unknown, interesting, and useful, spatial patterns and relationships within spatial datasets. SDMS interoperability refers to the ability of diverse SDM systems to cooperate towards global goals by communicating and exchanging spatial data, patterns, relationships and other intermediate results.

SDMS interoperability is important for the following reasons. Many spatial datasets are extremely large and distributed over multiple sites. It is often not possible to bring the entire spatial dataset to one location for mining interesting global patterns due to concerns like privacy, security, communication cost, etc. SDMS interoperability may be helpful in overcoming these barriers. For example, one may run individual SDM system at each site to extract local patterns from the subset of spatial data stored at each site. The local SDMS may take advantage of interoperability to interchange local patterns and intermediate results to extract global patterns.

SDMS interoperability poses many difficult challenges at multiple levels. At the highest level, semantic interoperability level, it needs a specification of agreement about content descriptions of spatial data, patterns and relationships. Ideally, a common vocabulary of concepts and a common ontology are shared among all systems. Alternatively, well-defined translations are available to map concepts used by the sender to those used by the receiver. The middle level may focus on structural interoperability to provides means for specifying semantic schemas (or meta data) for sharing. At the lowest level, syntactic interoperability specifies common message formats (e.g. tags and marking) to interchange spatial data, patterns and relationships.

Recent development in spatial web standards (e.g. Web Mapping Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS), Web Terrain Service (WTS), Geographic Markup Language (GML), etc.) have addressed many issues in syntactic and structural interoperability related to spatial data. However, these standards have not directly addressed exchange of spatial patterns and relationships. One may encode spatial patterns as spatial data to establish basic communication among cooperating SDMS. However, it will still leave a fair amount of work towards creating common data mining concepts, ontologies etc.

Another option is to use a combination of spatial web standards and web based data mining standards (e.g. Predictive Model Markup Language (PMML)). PMML supports syntactic and structural interoperability for classical data mining by providing facilities like data dictionary, mining schema, transformation dictionary, model statistics and model parameters. However, they focus on classical data mining patterns, e.g. regression, which assumes that learning data samples are independent from each other. This assumption rarely holds on spatial datasets, which exhibit high auto-correlation. SDMS has developed new concepts (e.g. colocation [3] , spatial outliers [4], location prediction[2]) and models (e.g. spatial auto-regression, join-based colocation [3] ) to address these limitations of classical data mining.

Thus, it is important to extend spatial web standards and web based data mining standards (e.g. PMML) to support novel SDMS concepts and models. At semantic level, this requires development of a consensus among SDMS researcher, users, and software developers towards a common set of concepts and ontologies. At structural level, it requires development of a consensus representation for exchanging schema. At syntactic level, it requires development of common message formats (e.g. XML tags).

Other key challenges include development of distributed algorithms for mining spatial patterns (e.g. colocations, spatial outliers, parameters of spatial auto-regression models) to compute global patterns from local patterns without copying local datasets to a common site. For example, our recent work [4] showed that almost all tests for spatial outliers belong to a special subclass of statistical functions, namely algebraic functions, which can be decomposed easily. This result provides a foundation for developing distributed algorithms for detecting spatial outliers by exchanging a few intermediate aggregate results without requiring exchange of the large datasets. We are now working on similar algorithms for mining other spatial patterns and relationships.

## REFERENCES

1. S. Shekhar and S. Chawla, Spatial Databases: A Tour, Prentice Hall, 2003 (isbn 013-017480-7.
2. S. Shekhar, P. Schrater, W. Vatsavai, S. Chawla, W. Wu, Spatial Contextual Classification and Prediction Models for Mining Geospatial Data, IEEE Transactions on Multimedia, 2(4), June 2002.
3. S. Shekhar, Y. Huang, Discovering Spatial Colocation Patterns, Intl. Symposium on Spatial and Temporal Databases, 2001. (Extended version to appear in IEEE Transactions on Knowledge and Data Engineering).
4. S. Shekhar, C. T. Lu, P. Zhang, A Unified Approach to Spatial Outlier Detection, GeoInformatica: An Intl. Jr. on Adv. in Computer Science for GIS, 7(2), 2003. (A summary of results appeared in ACM SIGKDD Intl. Conf. on Data Mining an Knowledge Discovery, 2001).

# GEOSENSOR NETWORKS AND SPATIAL WEB

***(Position Paper)***

**ANTHONY STEFANIDIS**

Dept. of Spatial Information Science & Engineering
National Center for Geographic Information and Analysis
University of Maine

## GEOSENSOR NETWORKS (GSN)

A geosensor network (GSN) can be loosely defined as a sensor network that monitors phenomena in geographic space, and in which the geospatial content of the information collected, aggregated, analyzed, and monitored is of fundamental importance [Stefanidis & Nittel, 2004]. Analysis and aggregation may be performed locally in real-time by the sensor nodes or between sensor nodes, or off-line in several distributed, in-situ or centralized repositories. Regardless of where these processes take place, the spatial aspect is dominant in one or both of the following levels:

- *Content level,* as it may be the dominant content of the information collected by the sensors (e.g. sensors recording the movement or deformation of objects), or
- *Analysis level,* as the spatial distribution of sensors may provide the integrative layer to support the analysis of the collected information (e.g. analyzing the spatial distribution of chemical leak feeds to determine the extent and source of a contamination).

The geographic space covered by the sensor network, or analyzed through its measurements, may range in scale from the confined environment of a room to the highly complex dynamics of an ecosystem region.

The use of sensor networks for geospatial applications is not really new. Satellites and aerial cameras have been providing periodic coverage of the earth during the last few decades. However, the evolution of sensing devices [Hellerstein et al., 2003; Howard et al., 2003] is revolutionizing geospatial applications. The old paradigm of calibrated sensors collecting information in a highly-controlled deployment strategies is now substituted by wireless networks of diverse sensors. This evolution has a profound effect on the nature of collected datasets:

- Homogeneous collections of data (e.g. collections of imagery) are now substituted by heterogeneous feeds for an area of interest (e.g. video *and* temperature feeds).
- Regularly sampled datasets (e.g. coordinates of similar accuracy in a regular grid) are substituted by pieces of information that vary substantially in content, resolution, and accuracy (e.g. feeds from few distinct irregularly distributed locations with sensors of varying accuracy).
- Information becomes increasingly spatiotemporal instead of just spatial, as sensor feeds capture the evolution over time of the properties they monitor.

This evolution is bringing forward substantial challenges in terms of data management and analysis, but at the same time introduces up-to-date, unparalleled scene modeling capabilities. Information can be collected in real-time, allowing the monitoring of phenomena and events, and enabling the continuous updating of geospatial databases.

## GEOSENSOR NETWORKS AS SPATIAL MINIWEBS

Based on the above, GeoSensor Networks may be viewed as Spatial MiniWebs. They are spatially-focused information communities that actively share geospatial data over a limited area, with nodes collecting, forwarding, and even analyzing and aggregating information. Extending these capabilities is a major challenge facing GSNs, namely distributing analysis and aggregation capabilities across the network in order to optimize its functionality and performance. At the same time, an emerging challenge is to get multiple GSNs to collaborate, thus extending their spatial and thematic coverage.

These challenges put renewed focus to the typical issues addressed in spatial webs (e.g. interoperability, accuracy management), while imposing additional constraints like:

- the need to extend current modeling capabilities (e.g. Virtual Reality models) to handle the near real-time information flow of geosensor networks,
- the importance of distance in analysis, an underrepresented issue in today's spatial web, and
- computational cost, as network nodes typically have limited capabilities and/or energy resources.

## REFERENCES

Hellerstein J., W. Hong, and S. Madden, 2003. The Sensor Spectrum: Technology, Trends, and Requirements, *SIGMOD Record,* 32(4), pp. 22-27

Howard A., M. Mataric, and G. Sukhatme, 2003. From Mobile Robot Teams to Sensor/Actuator Networks: The Promise and Perils of Mobility, downloadable from www.spatial.maine.edu/~gsn03/program.html

A. Stefanidis A. & S. Nittel (eds.), 2004. *GeoSensor Networks,* CRC Press

*Position paper for Specialist Meeting on Spatial Webs, NCGIA, December 2-4, 2004*

**Developing Intelligent Software Agents for Semantic Interoperability on Internet GIS.**

Author: Ming-Hsiang (Ming) Tsou, Department of Geography, San Diego State University

The progress of geospatial information technology and the need for global distribution of geographic information is pushing the GIS community to develop Geographic Information Services (GIServices) on the Internet. Currently, on-line GIServices are well adopted in academic research projects, federal and local governments, and the GIS industry. Data sharing is now commonplace, but the lack of semantic interoperability becomes a major hurdle to advance the development of distributed GIServices. This paper describes a software agent-based communication mechanism (GeoAgents) that dynamically integrates geospatial data and GIS services across the Internet. The goal of Internet GeoAgents is to reduce user work and information overload by interpreting, filtering, and converting semantic information automatically. The adoption of intelligent software agent technology can facilitate the semantic interoperability on Internet GIS applications. Three fundamental roles for GeoAgents are essential to distributing GIServices: information broker, information interpreter, and decision maker.

An information broker helps users search requested information, and filter out unnecessary elements according to user-supplied specification. The interaction process between software agents and users is similar to the relationship between travel agents and airplane travelers. When a traveler tells the travel agent their preference, i.e. their dates of travel and the names of departure and arrival cities (a task specification), the travel agent filters and simplifies information from complicated flight schedules to create a few possible itineraries for the traveler to choose from. The travel agent will in theory only provide a reasonable range of choices to the buyer. Similarly, a software agent (information broker) will filter out unnecessary information and provide a reasonable number of choices to the user.

An information interpreter agent can access and convey information. In distributed network environments, heterogeneous data models and systems cannot communicate directly. Geospatial data present special challenges as they are characterized by spatial dependence, spatial heterogeneity, and sensitivities to spatial and temporal resolution. These properties complicate data fusion and transfer protocols. A GeoAgent can bridges heterogeneous information islands to convert between different data types and models. To translate the information correctly, a GeoAgent must acquire related GIS ontology and semantic translation methods. The knowledge and methods can be previously defined and encapsulated in the metadata of object-based components or be provided by external ontology databases/entities.

A decision maker agent collects and analyzes information according to specific events, makes an optimal decision in collaboration with a user or other agents, and executes a behavior accordingly. The behavior of decision maker agents is based on the rules defined by its own knowledge base, by user-defined rules, or by the suggestions from other software agents. Decision maker agents require appropriate user interfaces for accepting decision-making rules and for procedure verification. Also, appropriate communication protocols must be formalized to enable message passing for agent interaction. Another issue involves how an agent chooses collaborative agents for the decision making process. Since a decision maker agent will have more power than the other types of agents, a hierarchy of software agents must be defined. For example, in a network security system, a certificate authority (CA) server (higher level) can assign or distribute public keys or private keys to the client web servers (lower level) for improving the security of web browsing and transactions. To obtain the certificates, the client web servers need to send their requests to the CAs. The CAs (decision makers) then decides whether to issue the certificates to the client servers or not.

The advantage of adopting software agents is that complicated GIS tasks can be executed under software agent control to provide more comprehensive services for information access, retrieval and processing. Consider the following request: *Finding GIS data [Roads] in [San Diego]*. In a current Web search engine process, the search can only have two possible answers: Found, or Not-found. In an intelligent software agent framework, software agents can create more different answers as the following:

*Request* *Software Agent Response*
(Find, [Roads, San Diego]) → (Please re-define the keyword "San Diego" for 1: City 2. County)
→ (Search Result A: Location = Web site: GIS.SDSU.EDU)
→ (Alternative choice: replacing [Roads] by [ Major Roads], then search)
→ (Alternative choice: selection [Roads, California] from Web-B with additional GIS process required (Clipping [Boundary-of-San-Diego] from California)

One important criterion for the implementation of intelligent software agent framework is the establishment of GIS ontology. An ontology is a particular conceptualization of a set of objects, concepts, and other entities about which knowledge is expressed, and of the relationships among them. GIS ontology defines knowledge that guides GeoAgents' behavior. The following statements provide a few example statements of GIS ontology.

- *The map units in State Plane Coordinate System are feet.*
- *1 feet = 0.3048 meters.*
- *'Roads' include 'major roads, highways, unpaved roads'.*

By adopting these rules, a software agent can make inferences to accomplish tasks. For example, if a geodata agent receives a request for "buffer at 200 feet" on an ArcView Shapefile with UTM coordinate system (meter-based). The geodata can retrieve these rules from the ontology and convert 200 feet to 60.96 meters, and revise the buffer operation accordingly. To demonstrate the feasibility of GeoAgents, a Java-based prototype of an intelligent data processing system for environmental models was developed to help end users search and retrieve geographic information from distributed and heterogeneous data sources via the Internet. The adoption of GeoAgents in the prototype can automatically convert the retrieved data to readable formats required by specific environmental models.

To summarize, "sharing" is the most essential motivation of distributed GIServices. To sharing data, information, and knowledge among the GIS community, we need to develop a flexible framework to exchange semantic information across networks. Adopting intelligent software agents has great potential to facilitate such framework for semantic interoperability.

**Nancy Wiegand**

# Position Paper for the Specialist Meeting on Spatial Webs

## Internet DBMSs and Semantic Integration

The perspective presented here is based on Internet Database Management Systems (DBMSs) and semantic interoperability relevant to querying and locating geospatial data. This work is a combination of geospatial and DBMS technologies. A new vision is presented to query geospatial data distributed across the Internet. That is, DBMS technologies are being advanced to enable full DBMS-type querying over the Web without data being resident in a DBMS on a particular server. This emerging technology can be considered to be a great extension compared to HTML search engines, which provide searching over the entire Web but only allow keyword searches and are restricted to returning URLs and not actual answers. This new technology can also extend the functionality currently found in geospatial clearinghouses/portals, which restrict users' keywords to pre-defined categories and only return URLs of data sources.

Contrary to this, an Internet DBMS allows full query expressions and the return of answers. An example is to find the *names* of land owners whose property is classified as cropland and who have property larger than the average size property. To obtain this information without an Internet DBMS, a user would have to locate the appropriate data set, download it, and put it into a local GIS or DBMS. However, Internet DBMSs will allow the user to easily pose many ad hoc queries and perform analyses that are now cumbersome. Over the long term, the technology has promise to promote the dissemination of information and derived knowledge to society. Niagara [NDM+01] is an example prototype Internet DBMS. It has search and query engines to locate data sources and query over them. It indexes data marked up in XML, and data are queried using an XML query language. For example, XQuery is being developed by the W3C as a standard XML query language.

## Current Prototype System

We have been studying the use of Internet DBMS technology for geospatial data and, in particular, its use for a Wisconsin Land Information System. We are also extending our work to apply to other geospatial portals such as Geospatial One-Stop. We developed a prototype system that extends the Niagara architecture to illustrate our approach [WZ04, WZC+03, WZC+02]. For example, to find data sources relevant to a query that covers a geographic area of more than one jurisdiction (e.g., find all cropland in a watershed), we extract minimal metadata from the user through our geospatial interface. We also extract query terms based on an ontology for a theme. Our global ontologies are supersets of values to avoid contentious issues in defining a standard. Our ontology subsystem performs look-ups to rewrite generic *GeoSpace* queries in local terms for each data set. We are working on the most heterogeneous type of data, which is to resolve differences in land use coding systems. To do this, we extended work on schema and semantic integration to the *value* level of attributes.

Semantic interoperability is a challenge, and we experimented with several methods. Our first and, ultimately, most accurate method was to have a domain expert perform mappings from a global ontology to each local ontology. Although we started with a manual approach, this task was greatly enhanced by the use of a tool [CRS+02]. The global to local mappings are stored in look-up files (agreement files) in XML format that are stored locally but indexed centrally. To alleviate the manual decision-making, we also experimented with adjustments to the Naives Bayesian classifier [Zho03] and with Formal Concept Analysis to try to automatically match global to local ontologies [ZW]. We hope to do more work with these methods to fully realize their potential and limitations. Our partner in our NSF Digital Government project is also working on other methods for semi-automatic deductions [CSC04].

Our ongoing work will combine and continue to explore methods that more closely approach an automatic resolution of query terms. Our work can also be moved into the new technologies of RDF and OWL that were not available when we started. However, we found that our semantic expressions cannot be stated fully enough in OWL, and we hope to make contributions there. We are also working on more efficient ways to extend Internet DBMS architectures to accommodate an ontology subsystem with look-up files, including investigating P2P architectures. In addition, we are considering more extensible indexing methods for efficient access to large amounts of semantic look-up data.

**Locating Geospatial Data Sources Automatically**

In addition to the above work, we have proposed the use of Internet DBMS technology to help solve the problem of *locating* geospatial data [Wie04]. That is, current methods often involve searching using HTML search engines, the effectiveness of which we already noted could be greatly enhanced with DBMS technology. Other methods involve the data publisher registering with one to many geospatial Web sites and the user visiting those sites to find data. However, many sites exist (FGDC, ADL, the Geography Network, state and local sites, etc.) making it difficult to know which site to publish in or visit. Although Geospatial One-Stop may alleviate this problem, it could be some time before all data are available in one place. Also, other types of methods have been proposed for handling the dissemination of geospatial data [e.g., OCC+04].

The proposed solution here is to encourage geospatial data producers to publish their FGDC or other metadata files in XML over the Web along with, but separate from, the source data files. Then, Internet DBMS technology would allow querying of the metadata files to specifically locate a required metadata file, which would contain the URL of the source file. In this manner, source files can be precisely located. Furthermore, because many geospatial applications involve the same types of data sets with the only variant being the particular geographic area or jurisdiction, a Web application could be developed to automatically locate data sources. That is, a Web application or service for land use planning (or emergency response) would contain query templates to locate various types of data sources (land use, wetlands, roads, etc.). Template criteria would range over various metadata fields. The Web service would automatically substitute variable information (e.g., jurisdiction name) into a query template and send it to an Internet query engine, which would process the query and return the URL. Because metadata formats are not all the same, semantic technology is needed to do look-ups between possible metadata attributes and values and rewrite queries until the appropriate data source can be found. Such query rewriting techniques would be similar to our existing prototype geospatial Internet query system.

[CSC04] Cruz, I.F., Sunna, W., and Chaudhry, A. 2004. "Semi-Automatic Ontology Alignment for Geospatial Data Integration", In M. Egenhofer, C. Freksa, & H. Miller (Eds.) *Proceedings GIScience,* Springer, pp. 51-66.

[CRS+02] Cruz, I. F., Rajendran, A., Sunna, W., & Wiegand, N. 2002. "Handling Semantic Heterogeneities Using Declarative Agreements", In A. Voisard, & S. Chen (Eds.), *Proceedings of ACM GIS*, pp. 168-174.

[NDM+01] Naughton, J., DeWitt, D., Maier, D., & others. 2001. "The Niagara Internet Query System", *IEEE Data Engineering Bulletin*, 24(2), 27-33.

[OCC+04] Onsrud, H, Camara, G., Campbell, J., Chakravarthy, N. 2004. "Public Commons of Geographic Data: Research and Development Challenges", In proceedings GIScience, Egenhofer, M., Freksa, C., & Miller, H. (Eds.), Springer, pp. 223-238.

[Wie04] Wiegand, N. 2004. "Querying Metadata Over the Web to Locate Geospatial Data", In Proceedings of GIScience 2004 Extended Abstracts and Poster Summaries, October, pp. 224-226.

[WZ04] Wiegand, N. and Zhou, N. "Ontology-Based Geospatial Web Query System", To appear as a chapter in a book resulting from the Next Generation Geospatial Information (NG2I) Workshop, Boston, MA, October 2003, Taylor & Francis.

[WZC03] Wiegand, N.; Zhou, N.; and Cruz, I.F. 2003. "A Web Query System for Heterogeneous Geospatial Data", In Proceedings Scientific and Statistical Database Management (SSDBM), July, 2003, pp. 262-265.

[WZC+02] Wiegand, N.; Zhou, N.; Cruz, I.; and Rajendran, A. 2002. "Querying Heterogeneous Land Use Data Over the Web", GIScience 2002 Abstracts, M. Egenhofer and D. Mark (Eds.), Sept. 2002, pp. 207-210.

[WZC+03] Wiegand, N.; Zhou, N.; Cruz, I. and Sunna, W. 2003. "Resolving Schema and Value Heterogeneities For XML Web Querying", Semantic Integration Workshop at the International Semantic Web Conference (ISWC), Sanibel Island, Florida, October 20, 2003, Demo paper, pp. 149-153.

[Zho03] Zhou, N. 2003. "Automatic Ontology Mapping of Categorical Information", *Proceedings of the National Conference on Digital Government Research*, Boston, MA, pp. 401-404.

[ZW] Zhou, N., & Wiegand, N. 2004. "Formal Concept Analysis for Semantic Integration", Unpublished.

**Position Paper submitted to *Specialist Meeting on Spatial Webs: Call for Participation***

# Syntactic Interoperability and Spatial Support through practical development

Chaowei (Phil) Yang, Assistant Professor
Earth Systems and GeoInformation Sciences
School of Computational Sciences, George Mason University
Fairfax, VA 22030-4444

The fast pace development of Geographic Information System (GIS) has benefited us in almost all aspects of our daily life by acting as a media (Sui and Goodchild, 2001). The development of distributed computing infrastructure calls to share geographic information resources to build inexpensive distributed GIS applications. The problems of incompatibility in format, syntax, semantics, quality, and spatial support emerged within the efforts of sharing the geographic information resources. Different standards organizations, such as OGC and FGDC, are trying to set up standards and implementing GISs based on relevant standards to solve the interoperability problems. In the context of Geographic Information Science (GIScience), Goodchild (2004) suggested to research on spatial webs and data integration to address them in a relevant NGA sponsored project.

We actively practice in solving interoperability issues through the project of *Virginia Access – Middle Atlantic Geospatial Information Consortium* (NASA NAG13-01009) and the project of *Building an Interoperable Web Mapping Portal for the Middle Atlantic Region* (USGS/FGDC CAN. 03HQAG0146).

1. We participated in developing the FGDC framework data content standard (http://www.fgdc.gov/RReview/RRlogin.php) for Digital Elevation Model and are in the process of reviewing the standards. The FGDC framework data content standard involves the differentiation of metadata for each of seven themes and five sub-themes of transportation. This experience gives us the opportunity to observe the theoretical aspect of metadata for syntactic interoperability.

2. Collaborated with NASA Geospatial Interoperability Office, we are mapping our interoperable efforts to *Geospatial Interoperability Reference Model*, which recommends different standards for different aspects and levels of services, for example, EPSG, OGC WMS, FGDC Metadata and OGC GML/WFS are adopted to comply with Geospatial Reference Systems, Maps and Visualization, Metadata and Catalog Access, and Data Access standards/specifications respectively.

3. FGDC clearinghouse node and Geospatial One-Stop harvested portal are also constructed to exchange geospatial products with government agencies, such as USGS, and general public using Z39.50 and FGDC metadata standards

(http://gis.scs.gmu.edu/metaweb/smms.asp). This research and development effort gives us opportunities to observe the implementation aspect of metadata for syntactic interoperability.

4. Our research has been focused on different issues of distributed GIS since 1997 (Yang et al, 2004c). The research resulted in 3 software prototypes, products. Data generalization and abstraction methods are adopted and implemented in our high performance WebGIS prototype (Yang et al, 2004a). This practical development, we believe can contribute to the effort of providing spatial support.

5. Advanced distributed geospatial information computing methods, such as GridGIS (Yang et al, 2004b), is also under research and development. We are trying to develop a suite of middleware services (Yang and Tao, 2004) to share geospatial information resources, including data and process, in the cyber-infrastructure as envisioned by NSF (http://www.nsf-middleware.org/) and Gehegan (2004). We hope to provide practical development contribution to the interoperability issues of GIScience and contribute to all the four topics identified by the meeting of *specialist meeting on spatial webs* through the research.

Although our research and development have potential contributions to the four topics of the meeting, we envision our contributions to the spatial webs as 1) to the syntactic interoperability from both theoretical and practical metadata aspects, and 2) to the spatial support through our development and research on high-performance distributed GIS.

## References

1. Gahegan M., 2004, Grid Computing, the Semantic Web, and Visualization: Cyber-Infrastructure and the Future of GIScience, keynote in GIScience'2004, University of Maryland Conference Center, October 20-23, 2004
2. Goodchild, M. F., 2004, Strategic Enhancement of NGA's Geographic Information Science Infrastructure, http://www.ncgia.ucsb.edu/projects/nga/Executive_Summary.htm, a proposal to NGA.
3. Sui, D. Z., and Goodchild, M. F., 2001, GIS as a media, *International Journal of Geographic Information Science*, **15**, 387-390
4. Yang C., Wong D., Yang R., and Kafatos M., Li Q., 2004a, Performance Improving Techniques in WebGIS, *International Journal of Geographic Information Sciences*, in press.
5. Yang C., Kafatos M., Wong D., Yang R., Cao Y., 2004b, GridGIS: A next generation GIS, CITSA 2004, Jul. 21-25, Orlando, FL, pp.22-27.
6. Yang C., Kafatos M., Wong D., Yang R., 2004c, Network GIS, Earth Systems and Remote Sensing, Edited by Menas Kafatos and John Qu, London: Springer, in press
7. Yang C., Tao V., 2004, Chap.7: Distributed Geospatial Information Services, *Frontiers of Geographic Information Technology*, Edited by Rana S. and Sharma J., London: Springer, in press.

# Appendix C

## Participant Presentations*

### Thursday, December 2

Introduction
[Goodchild](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_intro.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_intro.pdf

[Loomer](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/loomer.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/loomer.pdf

Plenary Session 1
[Raubal, Lutz](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/raubal_lutz.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/raubal_lutz.pdf

[Mark](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/mark.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/mark.pdf

Breakout Groups I
[Goodchild](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_breakout1.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_breakout1.pdf

Synthesis I
[Gahegan](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/gahegan_breakout1_summary.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/gahegan_breakout1_summary.pdf

### Friday, December 3

Plenary Session II
[Goodchild](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_accuracy.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_accuracy.pdf

[Kyriakidis](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/kyriakidis.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/kyriakidis.pdf

Breakout Groups II
[Goodchild](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_breakout2.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/good_breakout2.pdf

Plenary Session III
[Cruz](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/cruz.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/cruz.pdf

[Agouris](http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/agouris.pdf)
http://www.ncgia.ucsb.edu/projects/nga/docs/pdf/agouris.pdf

**Presentations are in content-protected Adobe Portable Document Format (pdf), requiring Adobe Acrobat 5.0 or greater for viewing. Contact individual authors with questions about additional content access.*

# Appendix D

## Research Project Overview

This appendix material is intended to provide an overview of our collaborative research project, "Strategic Enhancement of NGA's Geographic Information Science Infrastructure." Additional material is available through the NCGIA web page, http://www.ncgia.ucsb.edu.

Users of geospatial technology are both blessed and cursed by the growing availability of geospatial data, software, hardware, and services. With the exponentially growing volume of spatial data comes a mind-boggling array of resources for answering questions as well as a host of significant problems. These problems include incompatibilities between multiple data sources and competing formats; the inability to organize, combine, and search datasets with varying spatial resolution and extent; and the social barriers that inhibit the sharing of valuable geospatial data resources between communities. Although geographic information systems have been portrayed as a solution to many of these problems, spatial data users still spend a large proportion of their time, effort, and resources trying to overcome the barriers and problems listed above. In an effort to address these problems and others associated with data sharing and data interoperability, the NCGIA (an independent research consortium dedicated to basic research and education in geographic information science) began a multi-year funded research effort titled "Strategic Enhancement of NGA's Geographic Information Science Infrastructure." The three NCGIA member institutions (The University of California, Santa Barbara, The University of Buffalo, and The University of Maine) all participate in this research effort under the direction of Michael F. Goodchild, Chair of the NCGIA Executive Committee. Our research project is divided into two principal phases, first, spatial webs, and second, data integration. These two research topics address several high-priority items in the research agenda of the University Consortium for Geographic Information Science (UCGIS, http://ucgis.org), and are of significant interest to the geospatial intelligence community, which needs the capability for rapidly assembling data from multiple sources to answer questions and to add value to existing geospatial datasets. Through this research, we are working to remove existing impediments that prevent individuals, groups, or institutions from sharing and integration geospatial data. The two research themes of this project, spatial webs and data integration, have a variety of interesting challenges that we are addressing.

## Spatial webs

We define a spatial web as a geographically dispersed information community actively sharing geospatial data and other digital geospatial objects. We recognize that there is no universally accepted and adopted standard for data format, data syntax, spatial support, accuracy, or semantics. As such, the critical task is to develop methods, techniques, and systems that recognize and compensate for inherent differences. A spatial web system therefore must have the capability for rapid data ingestion, format translation, aggregation, interpolation, registration, and adjustment. Because of the large proportion of time spent reformatting, reprojecting, resampling, and registering spatial data in any large geospatial project, a spatial web concept is needed. We project that significant financial and human resources can be redirected from the mundane preliminary data assembly task toward more valuable and critical analytical activities as a spatial webs concept is adopted by the geospatial community.

The concept of a spatial web echoes the development of information sharing communities in other domains, where information is collected from a variety of different locations distributed across a network. The advantage in such as system is that data does not need to be stored redundantly in every location where there is a demand. The development of computer network protocols and systems over the past three decades has made such a distributed information sharing system possible, but a distributed geospatial web adds the advantage of allowing data to be stored in a location where it can be easily maintained by a steward, ground-truthed, and adjusted relative to other local data.

Many problems and unanswered fundamental questions exist that prevent a fully functional spatial web from being established. These problems and questions are a part of our research agenda, and include 1) syntax issues, 2) spatial support issues, 3) accuracy issues, and 4) semantic issues.

**Spatial webs: Syntax issues**

The history of GIS is replete with the use of proprietary formants, competing and overlapping format standards, and limited open formats. Vendors and organizations approach this problem by writing format translators where possible and by creating open exchange formats as needed. Global Geomatics' Mapfusion product, originally developed for the Canadian military, and Safe Software's Feature Manipulation Engine (FME) product suite are widely regarded as state of the art in providing interoperability between hundreds of geospatial data formats The Open Geospatial Consortium's programs and initiatives in interoperability also provide a valuable resource. One of the serious remaining problems in syntactic interoperability for spatial webs is the lack of metadata with appropriate content, readily accessible and machine readable, and tightly coupled to the data. Metadata are absent from many older geospatial data formats, existing only as notes on obscure Web pages or references in publications. Newer geospatial data formats tend to have associated metadata, but it is frequently only defined for simple file formats and not readily adapted to multi-table databases (such as ESRI's Geodatabase format), and is seldom available for the entire collection of data as well as the individual geospatial objects contained within the dataset.

**Spatial webs: Spatial support issues**

There are two principal conceptualizations for the Earth's surface: continuous fields and discrete objects. The discrete object conceptualization is more appropriate in settings where there is persistence, internal homogeneity, determinate boundaries, and integrity. This conceptualization is used for objects such as buildings and other similar structures that can be neatly modeling with points, lines, areas, and volumes in data structures. A continuous field conceptualization is more appropriate for phenomena that vary continuously over the Earth's surface, such as topography, weather and atmospheric conditions, soil, and vegetation cover.

Representing a continuous field in a digital medium involves some necessary discretization, including six general methods outlined by Longley *et al.* 2001): sampling at irregular points, sampling at a grid of points, averaging over a grid of cells, digitized contours, triangular irregular networks, and irregular tessellations. These six methods create datasets of points, rasters cells, raster cells, polylines, polygons (triangles), and polygons. We term these objects the *spatial support* of the continuous field representation. Interoperability problems arise in integrating data representing continuous fields with other fields or with data representing discrete objects, because in general the spatial support of the two data sets will be different.

## Spatial webs: Accuracy issues

One of the most problematic of impediments to interoperability is the tendency of geospatial data sets to exhibit varying levels of accuracy. U.S. Federal standards recognize five components of accuracy: positional accuracy, attribute accuracy, completeness, logical consistency, and lineage (Guptill and Morrison, 1995). An extensive literature on geospatial data accuracy, and the broader issue of uncertainty, has been published in the past 15 years. But while this research has addressed description, modeling, visualization, and error propagation (Zhang and Goodchild, 2002), there has been very little research on the issue of interoperability, or the degree to which two data sets of different levels of accuracy can be used simultaneously. Goodchild (2002) has discussed the issue of positional accuracy in this context, and has presented ways of achieving interoperability through what he terms *measurement-based* GIS, and Duckham (2001, 2002) has argued that object-oriented designs are particularly suitable for this purpose. Research is needed to address interoperability across different degrees of attribute accuracy and completeness, and to examine the impact of such interoperability on GIS queries, analysis, and inference.

## Spatial webs: Semantics issues

Geospatial semantics includes all elements related to conceptualization of geospatial data, including the definitions of geospatial entities, geospatial processes, and spatial relationships. Some current research foci in this area include the impact of human cognitive processes on geospatial semantics, the impact of human linguistic description on geospatial ontologies, and the semantics of technical symbols used in geospatial applications and data sharing communities. Geospatial data users often encounter difficultly in sharing data because of the way the geospatial data or geospatial processes are conceptualized. Semantic differences present a significant barrier in data sharing and interoperability.

A common problem in geospatial semantic interoperability is the use of distinct systems of classification, of soil condition for example, or of land cover type. Since conditions vary dramatically over the Earth's surface, it is inevitable that a classification scheme devised to fit the needs of one area will not match a scheme devised for the same purposes in another area (a vegetation cover classification scheme for the Amazon basin will be very different from one devised for Canada, for example). Semantic differences impact the definitions of attributes of a single feature class, and also impact the definitions of classes (the meaning of the class "road" will vary across the Earth's surface), and the feature types associated with each class ("road" may imply a polyline in some data sets, and a polygon in others). Semantic differences arise because of differences of language and script, and also because of differences of culture.

Over the past decade much research effort has gone into the conceptualization of the problem, and into technologies to address semantic issues. The term *ontology* has been employed to describe the generic issue, and also to cover the search for common semantics that can act as a central authority or source (Winter, 2001). For example, Weigand *et al.* (2003) describe efforts to achieve interoperability between the various counties of Wisconsin, each of which has adopted a distinct system of land use classification. A degree of interoperability has been achieved by devising an ontology, a state-wide common classification, that can be mapped to each county's distinct classification. In another version of this approach, ontology is conceptualized as a fundamental of human cognition, from which each culture, language group, or region obtains its terms by mapping, aggregation, or selection.

## Spatial Webs Research Agenda

Our research in spatial webs and data integration addresses each of the areas identified above. Of particular concern are the following issues:

- *Syntactic interoperability for complex multi-file databases*. We are investigating the formulation of database-level metadata, its relationship to table-level metadata, and its use to support interoperability. We are constructing prototypes using ESRI's Geodatabase as a case study, and use them to investigate the specific problems of syntactic interoperability that arise in this context.
- *Interoperability across differences of spatial support*. We are enumerating all of the types of issues that can arise in this area, and identifying solutions, by exploiting published methods and by devising new ones. We are integrating these methods into prototypes for transparent interoperability, and are demonstrating their use in a series of case studies.
- *Interoperability across accuracy differences*. In this area we focus on positional accuracy. We are devising and testing methods for achieving transparent interoperability between data sets of different accuracies. We are investigating the use of these methods in the context of a broad range of GIS functions, including overlay, buffering, and containment. We are formulating extensions to metadata specifications to address the use of such techniques, by exploring the concept of *interoperability metadata*, and defining the potential for data sets to be used together.

- *Semantic interoperability*. We are surveying all existing techniques in this area, and assembling them into a coherent framework. We are evaluating each technique in terms of performance, and also in terms of the complexity of the supported mappings. We are also nvestigating the integration of this type of interoperability with the three listed above. We are identifying suitable case studies, and building demonstration prototypes.

## Data Integration

The focus of our spatial webs research is achieving interoperability among data sets. The next logical step is data integration – to combine data sets into a single product. This can be done through concatenation, or the combination of multiple sources into a single merged source that contains all the input; through averaging, where multiple sources is combined through non-reversible arithmetic averaging into a single data source; and through differencing, where the creation of a single data source represents the difference of the inputs. We will address these three forms of data integration in a comprehensive manner on all common types of geospatial data. We propose the use of a more general terms, data conflation, to describe a comprehensive approach to the data integration problem that will address multiple data types and multiple purposes. Our data integration research, to take place following the spatial webs research, will focus on the following topics:

- *Identification of a complete set of conflation use cases*, within the formal structure of Unified Modeling Language (UML); construction of a theoretical framework that includes all geospatial data types and use cases; and a comprehensive assessment of the available and tested techniques within this framework.
- *Research on high-priority gaps in this assessment*, focusing on challenging problems that are of obvious value in the geospatial intelligence community, and building prototypes to test and demonstrate solutions using appropriate case studies.
- *Integration of the results of this research with those obtained under the spatial web initiative*, in order to create a comprehensive system for interoperability and integration of disparate geospatial data sets.

*Extension of the metadata work related to spatial webs*, to include information of relevance to integration, and the automated creation of metadata for integrated data sets.

# Appendix E

## References

Duckham, M., 2002. A user-oriented perspective of error-sensitive GIS development. *Transactions in GIS* 6(2): 179–194.

Duckham, M., 2001. Object calculus and the object-oriented analysis and design of an error-sensitive GIS. *GeoInformatica* 5(3): 261–289.

Fonseca, F. and Egenhofer, M. (1999) Ontology-Driven Geographic Information Systems. in: C. B. Medeiros, (Ed.) 7th ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO, pp. 14-19.

Goodchild, M.F., 2002. Measurement-based GIS. In W. Shi, P.F. Fisher, and M.F. Goodchild, editors, Spatial Data Quality. New York: Taylor and Francis, pp. 5–17.

Gruber, T.R. .A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199-220, 1993.

Guptill, S.C. and J.L. Morrison, 1995. *Elements of Spatial Data Quality*. New York: Elsevier.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 2001. *Geographical Information Systems and Science*. New York: Wiley.

Wiegand, N.; Zhou, N.; and Cruz, I.F., 2003. A Web Query System for Heterogeneous Geospatial Data", *Scientific and Statistical Database Management, SSDBM*, July, 2003, pp. 262-265.

Winter, S., editor, 2001. Special Issue: Ontology in the Geographic Domain. *International Journal of Geographical Information Science* 15(7).

Zhang, J.-X. and M.F. Goodchild, 2002. *Uncertainty in Geographical Information*. New York: Taylor and Francis.