FY 2007 Call for Proposals

U.S. Geological Survey, Geospatial Information Office

Center of Excellence for Geospatial Information Science (CEGIS) Research Prospectus

---

<u>Title:</u> GEOLEM: Improving the integration of geographic information in environmental modeling through semantic interoperability

<u>Principal Investigators:</u>

- Dr. Barbara Buttenfield, Professor, Dept. of Geography, University of Colorado – Boulder, Guggenheim 110; 260 UCB, Boulder, CO 80309-0260, (303)492-3618, babs@colorado.edu
- Dr. Olaf David, Computer Scientist – Colorado State University/USDA Agricultural Research Service, 2150 Centre Avenue, Building D, Suite 200, Fort Collins, CO 80526-8119, (970) 492-7312, olaf.david@ars.usda.gov
- Dr. Charles G. O'Hara, Research Professor, GeoResources Institute, Mississippi State University, 2 Research Blvd., ERC, Room 201, Starkville, MS, 39762, (662) 325-2067, cgohara@gri.msstate.edu
- Roland Viger, Physical Scientist – USGS, MS 412 Box 25046 DFC, Denver CO, 80225, 303/236-5030, rviger@usgs.gov

<u>Primary Contact:</u> Roland Viger

Project Description

The purpose of this research is to define an approach that will improve the discovery, evaluation, and appropriate exploitation of processing and database services for providing geographic information (*geo-services*). Current approaches for describing and selecting such geo-services fail to include computationally formalized, semantically relevant definitions for the information that they will produce. The term *definition*, as used here, is more than the metadata, that is, the technical characteristics commonly used to describe data. *Definitions* also must be able to include description of the phenomenon that the data represents, its thesaurus encoding, how it is represented, and how it relates to other phenomenon. The proposed research will define ways to semantically enrich standard forms of metadata commonly in use today. The effectiveness of this approach will be demonstrated by integrating geographic information systems (GIS) and a selected environmental model (EM). The significance of this demonstration is not only that it enables a GIS and an EM pairing to interoperate, but also that it facilitates interoperability through explicit semantic descriptors of shared information. The products of this research will improve interoperability of geographic information in general.

Background

Increased demands and concerns about limited environmental resources make the ability to manage natural resources ever more critical. Environmental modeling is one of the main tools for simulating and analyzing the behavior of these resources (Fedra 1993). GIS has long been advocated as a tool to help solve environmental problems (Dangermond et al. 1981). In particular, GIS has served as a powerful tool for the production of the geographic information used in environmental models. The integration of GIS, or more broadly, geo-services, with environmental models enables "detailed consideration of landscape properties and spatially distributed processes" (Goodchild, Parks et al. 1993). Ultimately, this integration is important because it can facilitate the improved management of earth resources. The use of geo-services to support modeling has become widespread as digital data, computing resources, the Internet, and GIS technology have improved and become easily accessible.

Generic methods for integrating GIS and environmental modeling have been difficult to develop because environmental simulation models and modeling frameworks fail to represent or describe geographic information in a rigorous way. Although models may define the units and characteristic a piece of geographic information refers to, they frequently lack a formalized description of what individual parameters actually *mean*, and how sets of parameters collectively *represent* features found in the landscape. While environmental simulation models are typically published with written documentation, the explicit discussion of geographic information in these documents is generally underdeveloped. This can create confusion for the model user as to the exact nature of the geographic information used within a model. The same holds true for users browsing Internet accessible geodata stores, such as the National Map. The interpretation of written documentation is typically the responsibility of the user of the model or data. This can lead to unintended, inconsistent, and potentially inappropriate usages. Such usages can easily lead to a reduction in the accuracy of results, whether they are model simulations or merely cartographic presentations.

In addition, there is a lack of description of a model's geographic information that is usable for automated computer processing. This omission prevents the model itself from warranting against inappropriate usage or driving the process of creating the parameters about a location's geographic characteristics. Further, the simplistic representations of geographic phenomena typically used by models are radically different from those found within GIS. Details about the correspondence between a model's representations and those found within GIS are not typically available. Details about how a GIS should be used to generate that information do not exist.

The uncertainty and lack of consistency concerning the meaning of geographic information and its treatment in modeling has contributed to the popularity of ad hoc approaches to integrating GIS and environmental modeling. These approaches tend to tightly integrate specific technological choices together, yielding solutions that cannot be generalized to other combinations of GIS or environmental models. The integration tends to intermix logic based on science with that based on the details of implementation, such as file formats and communication protocols. Once an integration has been carried out, the separation of meaningful information from the incidentals of the technologies leveraged is difficult. This intermingling has several negative impacts. Sharing the original geographic information, or that of the model, with other models is no easier as a result of the integration effort. The model cannot easily switch to use a different GIS. Integrating the GIS with another model has not been made easier. The details of the connection between the GIS and the environmental model are not easily visible. In short, the integration effort has not yielded any products that are readily reusable, or interoperable, in other contexts. Nor has it necessarily improved the understanding or documentation of the geographic information being used.

Hypotheses/Research Statements

1. The geoprocessing methodology used to generate the cartographic representation of a geographic feature supports a semantic definition of that feature and can function as an important, alternative form of knowledge about it. In the case of pre-existing data stores, meaningful knowledge can be dynamically generated based on a variety of contextual cues, such as language/terminology, usage patterns, and data formatting.
2. By formalizing a way to generically encode methodological and contextual metadata into a machine-readable form, domain-specific ontologies can be built and associated with a middleware infrastructure for handling ontological knowledge.
3. The coupling of ontological knowledge with a middleware infrastructure creates the ability to encapsulate non-semantically significant details of a geoservice; this promotes interoperability of geo-services by allowing a client to choose any semantically appropriate service provider. Further, this pairing can be used to carry comparative analyses of a data set's semantics.

Literature Review

This research draws on several bodies of literature. Beyond that which directly addresses types of integration of GIS and environmental modeling, an understanding of ontology, knowledge representation, semantics and interoperability are be used to architect and implement a generic knowledge handling system and concepts about interoperable geographic information. Using these ideas, in conjunction with object-oriented design principles, the proposed research will develop an approach for the generic specification of metadata about the geographic information produced by geoservice providers that can be used to carry out the integration of GIS and environmental models. The literature relating to the integration of GIS and environmental modeling, and that of object-oriented design will not be treated here due to page limits associated with this proposal.

Ontologies

An ontology is a hierarchical collection of all the types of representations and their associated definitions for a topical area, which is referred to as a *domain* or *universe of discourse*. An ontology might contain the word “dog” and associate a definition like, “has fur, four legs, and a cold nose,” with it. The ontology will not contain actual dogs (obviously) or references to specific, existing dogs. A *formal* ontology is a specification of literally everything. (e.g. http://wwww.cyc.com). A domain-specific or *material* ontology is used for more limited contexts (Peuquet, Smith et al. 1999). The formal specification of ontology provides a means of establishing the connection between the things within a domain and the domain itself (Guarino 1995). Gaurino (1995) states that this enables the representation of “an intelligent agent embedded in an external environment.” The development of ontologies for geographic information is an area of active research in the GIScience community (Frank 2001; Kuhn 2001; Pundt and Bishr 2002; Visser, Stuckenschmidt et al. 2002).

Ontologies provide *context* for conceptual information, as well as instances of those concepts. Individual concepts may be initially defined independently of each other, but the act of assembling them into an ontology explicitly defines the relationships between the individuals. Relationships may be established by simple separation within the ontology’s hierarchal graph, or semantic similarity of properties or behaviors. Context is potentially valuable information for the exercise of integration. If two domains (a GIS and an environmental model, for example) are being integrated, analogous concepts might exist within each. Even if the analogs are not identical, or even fully specified, the context surrounding a concept within one ontology might be sufficient to establish a semantic equivalence to its analog in the other ontology. Therefore, ontologies are useful not only for organizing information within an application, but as a database through which knowledge might be translated or brokered from one application context into another.

Kokla and Kavouras (2000) illustrated the generally useful ideas of semantic factoring and concept lattices. These promising techniques could help automate the mapping of semantics between ontologies. This mapping or translation of semantics is an important approach to interoperability, focusing on the equivalence of concepts rather than file structures. Visser, Stuckenschmidt et al. (2002) describe a semantic translator as a way to interoperate between ontologies for geographic information access and processing. Bishr (1998), using the term *schemata* in lieu of ontology, describes an approach to overcome at least certain types of heterogeneity across systems.

Knowledge Representation and Handling Systems

Leung (1997) states that a system of knowledge representation should consist of three parts. The first and most obvious part is a set of conceptual models, or *symbolic structures* in Leung's terminology, to express domain knowledge. The remaining two parts are a language and an inference mechanism. A language enables the system to reference and describe symbols and otherwise manipulate an ontology's knowledge content. Language itself is not a representation of knowledge, but a formalism for referring to and evaluating information. Logic is an example of a mathematically based language, which has syntax, semantics, and a system of reasoning. Reichgalt (1991) states that logic has become the backbone for modern knowledge representation systems. A language should provide at least the basic functionalities associated with first-order predicate logic, most important of which is the ability to form *propositions,* sentences containing a subject, a verb, and an object. The verb-object pairing is termed a *predicate*. The language should also enable the specification of material implications (i.e. if-then clauses), negation, and qualifiers such as *all* and *some*. Leung (1997) (p. 22) states that:

> The advantages of using first-order predicate logic for knowledge representation lie on the declarative, expressive, modular, and deductive nature of the logic...The ability to derive implicit knowledge from explicit knowledge…makes the formalism expressive. The modular representation of knowledge as a collection of predicates enables easy knowledge modification.

He goes on to state that the major drawback to using predicate logic as the language for a knowledge representation is its Boolean (i.e. binary) representation of truth. The ability to represent imprecision is a critical element in not only natural language systems, but also in human reasoning. This is often necessary because of incomplete knowledge, which is supported by general, contextual knowledge. Leung feels that fuzzy logic is a more "natural and general" way to represent knowledge.

Sowa (2000) states that knowledge representation systems must have a theory of reference to relate constants and variables to elements in the universe of discourse and a theory of truth to evaluate statements. The semantics associated with conceptual models create a system of meaning that is critical for evaluating the truth of statements about those conceptual models and instances of those conceptual models. This, combined with a system of reasoning, which specifies a set of rules by which complex combinations of simple expressions are evaluated, forms the *inference mechanism* named by Leung. Put another way, the inference mechanism of a language enables the derivation of implied information from explicitly known information. Sowa describes rules of inference that allow one pattern to be discriminated from another, while preserving the truth, as given by the language semantics. Most modern languages, including natural languages, enable inference.

Incorporating Semantics into Interoperability

It should be noted that the terms *integrated* and *interoperable* should not be assumed to indicate the same thing. Even if an integration technique is generic, it may not be a truly interoperable solution. Interoperability itself may have a variety of meanings, as discussed in Bishr (1998). This point is being reiterated here to emphasize that there are many ways that GIS can be used in conjunction with an environmental model. Different approaches have different strengths. Although none of the implementations mentioned below seem to develop a comprehensive approach for using semantics or the other ideas discussed above to carry out the generic and interoperable integration of GIS and environmental models (or another type of client), they provide valuable ideas that could be used to realize such a model.

Standards for interoperability, most notably for GIS, have not played a large role in efforts to integrate GIS and environmental models. The acceptance and active support of these standards by GIS vendors is still developing. For example, only one example was found in the literature documenting an effort to create interoperability for GIS using Open Geospatial Consortium (OGC) specifications (Wong and Swartz 2002), although this effort did not deal with the integration of GIS with environmental modeling. While other literature addressing this surely exists, the difficulty in locating it does indicate the relative dearth of this work in the GIScience community. The development by Wong and Swartz (2002) lacked any semantics pertaining to types of geographic information, instead focusing on the extraction of geometric features. This is due to the fact that specifications for this kind of activity are the most mature of those offered by the OGC. Specifications for the description of the semantics of geographic information are still evolving. The most promising OGC initiatives for this are the Geography Markup Language (GML) and Information Communities.

There are a variety of efforts that focus on interoperability for environmental simulation models. Several organizations that have formed within the environmental simulation modeling community include the HarmonIT Consortium (www.harmonit.org) in Europe, the Simulation Interoperability Standards Organization (SISO; www.sisostds.org), and the Interagency Steering Committee on Multi-media Environmental Models (www.iscmem.org) in the United States. The Commonwealth Scientific & Industrial Research Organisation (CSIRO; www.csiro.au) in Australia is not an ad hoc agency formed to investigate ways for creating cooperation between environmental models, but is very active in this field. All three of these organizations are involved in using semantics as a way to improve the interoperability of modeling software. SISO is promoting the use of High-Level Architecture to create libraries of geographic information types that encapsulate lower level spatial data models, presenting semantically richer ("higher level") views of geographic information to environmental models (Bernard and Krueger 2000). Bernard and Krueger (2000) establish abstract specifications for a "virtual GIS" application programming interface, and separately develop semantically higher-level specifications for data and functionality relevant to a handful of environmental models based on their virtual GIS. HarmonIT and CSIRO are also attempting to use generic conceptual models and high level system architectures to integrate different kinds of modeling (Abel, Kilby et al. 1994; Gijsbers, Moore et al. 2002).

Some of the most rigorous efforts to create interoperability rely heavily on database theory. Sheth and Larson (1990), in their paper "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases", present a general approach that has been widely referenced in the GIScience literature (Abel, Yap et al. 1992; Abel, Kilby et al. 1994; Lee, Madnick et al. 1996; Abel, Ooi et al. 1998; Bishr, Pundt et al. 1999; Rahm and Bernstein 2001; Ding and Foo 2002; Fonseca, Egenhofer et al. 2002). Abel and his colleagues have perhaps most seriously applied these ideas to what they refer to as the "systems integration problem." Their work is based on a three schema approach. They isolate external, conceptual, and internal schemas. The external schema defines the "look" of a component to external components. This is essentially an application-programming interface. The conceptual schema is analogous to a knowledge base of the conceptual model of the components. The internal schema is the specification data models, programming language, etc. that is used to implement the conceptual schema. Abel (1994) provide a clear overview of sources of schematic heterogeneity, which are barriers to integration.

Other approaches to achieving semantic interoperability apply the idea of a middleware to broker information from one component to another. Lee, Madnick et al. (1996) outline a technique to achieve interoperability between a source and consumer of information through the use of a *context mediator*. Within the GIScience literature, Visser, Stuckenschmidt et al. (2002) use a *semantic translator*, which is described as an agent that uses semantic information for the intelligent retrieval of information. They even discuss the possible extension of their semantic system to specify processing logic.

Ram, Park et al. (1999) present a semantic model for supporting GIS, and use it to realize a simulation modeling environment. Their contribution is heavily dependent on object oriented programming, and tends to avoid the discussion of this system's interoperability with components that are external to it. It appears that the type of interoperability that they develop is somewhat limited, although their approach is powerful. Robinson and Mackay (1995) also describe a semantic modeling environment for carrying out simulation. Their contribution does not attempt to address interoperability issues, but is an excellent example of a knowledge representation and handling system. Perhaps the most significant contribution of their effort is the development of a language and inference engine for forming predicates and propositions. They develop a variety of logical relation operators for geographic features. These two efforts are relatively GIS-centric, where those of Abel and Visser are more oriented towards the creation of a middleware.

A third approach to integrating GIS and environmental modeling is more model-centric. There are several examples of this. Villa and Costanza (2000) provide a relatively low level contribution with a Simulation Network Interface, which is essentially an architecture for specifying client-server relations between different environmental models, and potentially GIS. Jankowski (1992) details a higher-level approach to managing modeling, relying on knowledge-based methods and hierarchical decomposition to manipulate modeling concepts. Both of these approaches handle information and integration, but are not specifically geared towards geographic information or GIS.

Most of these approaches deal in pre-existing concepts and data. They use the semantics of these concepts of as a way to manipulate data. Gahegan (1996) takes a slightly more basic approach, in that he uses the means by which an instance of a geographic concept is created as the semantic of that concept. He states, "any meaning inherent within a dataset is intrinsically connected to the model by which it was captured." (p. 137). He later uses descriptions of the transformation of raw data into high-level geographic objects as the basis for establishing semantic equivalence for data stored in different forms (Gahegan and Flack 1999). Ram, Park, et al. (1999) perhaps presented work that was the most similar to that of Gahegan, but they did not focus on the use of semantics to enable interoperability.

The preceding literature review provides the theoretical basis for creating conceptual models (as opposed to data models) that are shared between GIS and environmental models. Instead of using general principles relating to organizing ontological hierarchies and creating knowledge-representation systems to create new GIS platforms, or interoperability focused exclusively on GIS or environmental models, these ideas can be used as ways to map information from the context of one pre-existing component to another. This research could therefore be readily extended and applied to many geodata providers, such as the National Map.

Objectives/Approach

The primary objective of this research is to improve a user's ability to access appropriate geographic processing and data. To do this, this research will explicitly encode knowledge ("semantic descriptors") about the information being sought into a new form of metadata and build a knowledge-handling middleware to broker information based on this metadata. These developments will provide a more semantically meaningful interface to these geographic information resources, thus improving the usability (and interoperability) of these resources.

This research seeks to also exploit these developments as a generic method for integrating GIS and environmental models. Because this integration is not specific to a particular pairing of a client and geo-service provider, the client has the ability to interact with alternative geo-service providers with minimal impact on the client. Further, this pairing of semantic metadata and middleware promote another objective of this research, which is the interoperability of geographic information across different clients. A second client could, for example, query the first client through the middleware as to the nature of its information. Based on what the middleware resolves from the semantic descriptors, the second client could make decisions about whether that information would be appropriate for its own use. In other words, by defining semantic metadata, we can improve our ability to integrate different kinds of science, which is a major thrust for the USGS.

The semantic interoperability developed here will focus on capturing the *geoprocessing methodology* associated with creating a GIS-based representation of a client's geographic information needs as a means of integrating it with a GIS. Geoprocessing methodology refers to the specific GIS operations performed on spatial data in order to generate a piece of geographic information that will be used a client. By encoding a description of geoprocessing methodology, the client can formalize ideas about a geographic information concept that it uses. This would, to a large degree, relieve the users of a client (e.g. an environmental model) of the burden of developing their own methodologies.

In addition to establishing more meaningful definitions of geographic information, a secondary purpose of this research is to develop methods for evaluating and comparing these definitions. Given a large number of geo-services, both automated and heuristically driven integration efforts face the difficulty of differentiating among these services to determine which are the most appropriate for a given task. As the number of geographic database (*geodata*) and especially internet-accessible geographic processing (*geoprocessing*) providers increases, the problem of "finding a needle in a haystack" will only grow. Providing ways to formalize how a data consumer can rationalize about these geo-services will be a significant contribution because it will improve the ability to understand what disparate geo-services actually provide. This has the potential to ultimately improve the scientific and operational validity of natural resource management.

Approach

Broadly, a watershed model will be analyzed and an ontology of geographic information defined for it. A geoprocessing methodology will be defined for each geographic information concept, and associated with it. The knowledge-handling middleware will be designed and built so that the watershed model can ask the middleware to produce instances of the geographic information concepts defined in its ontology. The middleware, in turn, will breakdown the watershed model's high-level requests for information into basic GIS commands and communicate these to a GIS server that will actually execute the commands. When the GIS server has finished, the middleware will extract the relevant information and return it to the watershed model.

The team will first analyze the USGS PRMS watershed model and define a material ontology for it. The ontology will be encoded into an XML format, using the terminology of the PRMS model. Then one (or more) geoprocessing methodologies will be designed for each geographic information concept within that ontology. Generic meta-specifications of these geoprocessing methodologies will be implemented in Java. References to the specifications will be added to the XML encoding of the PRMS ontology, forming the PRMS knowledge-base about geographic information. The PRMS model, running from within the USDA Object Modeling System, will then request information from the middleware using the terminology of PRMS. The middleware is GEOLEM (Viger 2004). GEOLEM will then consult with the PRMS knowledge-base and resolve what geoprocessing methodology is associated with the PRMS terminology. GEOLEM will make a connection to an actual GIS (in this case ArcGIS 9.2 via the python geoprocessor interface) and invoke the execution of the methodology. Only the final data product is returned to GEOLEM. All intermediate spatial data products are persisted exclusively within the GIS. GEOLEM will complete the transaction with the PRMS model by returning the requested information. This process will repeat with each request from the PRMS to GEOLEM. A working prototype of GEOLEM already exists and PRMS is well known to the model.

In order to demonstrate the interoperability of this approach with different GIS platforms, this entire sequence will be repeated with a different GIS server technology. We are currently evaluating GRASS and TerraLib as alternates to the original ESRI technology. The alternate solution will rely on the exact same XML-encoded ontology and the exact same Java meta-specifcations of geoprocessing methodology. Only the GIS server will be different.

In the second phase of the project, the team will develop techniques for browsing the set of meta-specifications of geoprocessing methodologies. This work will focus on using the methodological specification as a semantic descriptor, against which queries can be evaluated. Examples of the types of queries that we would like to be able to process are "rank all methodologies in terms of similarity to methodology X." or "Group all the methodologies into 5 clusters." Both of these examples are essentially spatialization problems, where an information space is constructed using the list of all possible commands/functions found in the library of meta-specifications that we wrote in support of PRMS. Each meta-specification can be plotted in this space based of the presence or absence of each possible command within the specification. The relative positions of meta-specifications within the information space will then be examined for suitability as an indicator of similarity between meta-specifications. In other words, we want to test whether similar geoprocessing methodologies yield semantically similar information. There are a large variety of ways to maninpulate the information space (e.g. dimension reduction techniques such as SOMs) that we would like to experiment with.

Facilities and Expertise: The group of investigators associated with this proposal is well qualified to successfully carry out the proposed research. They have a broad range of strengths including knowledge and experience in earth sciences, simulation modeling, programming, and geographical information systems and science. The diversity of experiences of this group will provide strong support for this cross-cutting research topic. With initial support from the EPA and ISCMEM, Roland Viger and Olaf David developed a working prototype of the GEOLEM knowledge-handling middleware, using Java, XML, and ArcInfo. This system has been documented in a Master's thesis (Viger 2004 The current proposal will take this work as a starting point and grow it to an operational system and, in the process, evolving concepts about the contents of semantic metadata and how to it can be used. Olaf David is the project chief and chief computer scientist for the USDA Object Modeling System (OMS) and is planning to use GEOLEM as the core infrastructure for processing and access to geographic information within OMS. Barbara Buttenfield is a well-established GIScientist who will bring a theoretical perspective to the group's efforts. Charles O'Hara brings expertise in the use of spatial analysis and image-based data streams to a wide variety of natural resource management issues. Jeff Hamerlinck is GIScientist who effectively serves as the geographic information coordinator for the state of Wyoming, and has research interests in geospatial infrastructure. Roland Viger has 10 years of experience in working with the PRMS watershed model.

Expected Results/Products

Technology Products:

1. A working implementation of the GEOLEM system (the middleware infrastructure), a sample application, and published documentation published.
2. A coupling of the GEOLEM system with the USGS PRMS rainfall-runoff model. This will include the development of knowledge-bases and specifications of geoprocessing methodology for PRMS that can be reused in other PRMS modeling efforts. Funds to support this will be externally secured from the NRCS.

Technology Transfer:

3. Placement of GEOLEM into an operational setting within the NRCS National Weather and Climate Center (NWCC). Funds to support this will be externally secured from the NRCS.
4. 3-day workshop to train NWCC employees on the use of GEOLEM. Funds to support this will be externally secured from the NRCS.
5. Inclusion of the GEOLEM core and applications within the USDA Collaborative Software Development Laboratory (https://colab.sc.egov.usda.gov/).

Information Transfer:

6. 1 journal article submission documenting GEOLEM concepts about the metadata-based specification of geoprocessing methodologies. The International Journal of Geographic Information Science and GeoInformatica are envisioned as possible target journals, Computers and Geosciences is another possible target.
7. 1 journal article submission to JAWRA describing the purpose of GEOLEM within the NWCC environmental modeling and decision support framework.
8. 2 conference presentations after the project is completed [COSIT 2007; GIScience 2008].

Significance to the USGS Mission

This research is relevant to USGS Mission because it will improve our fundamental ability to provide reliable scientific information to describe and understand the Earth. It will focus on the development of a set of approaches to significantly enrich standard FGDC metadata content that is typically associated with the science (e.g. models and data) we are responsible for providing to the public. Further, it will develop ideas on how to query and rationalize about data based on these augmented forms of metadata. These contributions will help to overcome the current semantic gap between the information that a user wants and the data content that we serve.

GEOLEM is an experiment in trying to find different kinds of semantically relevant metadata and using them in a geospatial framework to integrate information across different contexts. This has the potential to overcome issues of alternate physical representations (i.e. encoding within data structures) and focus on the true meaning of a data set’s content. GEOLEM will enable the explicit construction of ontologies describing geographic information. Using GEOLEM as a computational platform, new lines of research into rationalizing about geographic ontologies can be carried out. The research will also promote the interoperability of both geoprocessing and geodata services, and the interoperability of geographic information across different client usages.

One major impact that this research could have is to provide an entirely new approach to searching, understanding, and selecting geographic data sets from the ever-growing list of digital geographic data repositories of the National Spatial Data Infrastructure, such as the National Map. It seems difficult to overstate the value of being able to issue an NSDI-wide query for “streams or something like them” and have the query processing integrate understanding of what this means as opposed to simply matching character strings. Further research into dynamically generating and analyzing contextual, semantic metadata for static data stores an important topic for the next phase of this research.

There are a variety of broader impacts that this research will have. It will improve our ability to support other Federal agencies with regards to basic and applied science by streamlining the delivery and use of products such as environmental models and facilitating their integration with other scientific and management components of their decision-making infrastructures. The creation of this kind of semantic metadata will mean more explicit documentation of our science, which promotes communication, research, and improved understanding among academe and natural resource managers. Ultimately, this will yield improved science.

Selected References

Abel, D. J., P. J. Kilby, et al. (1994). "The Systems Integration Problem." International Journal of Geographical Information Systems **8**(1): 1-12.

Bishr, Y. (1998). "Overcoming the semantic and other barriers to GIS interoperability." International Journal of Geographical Information Science **12**(4): 299-314.

Bishr, Y. A., H. Pundt, et al. (1999). Probing the concept of information communities - a first step toward semantic interoperability. Interoperating Geographic Information Systems. M. Goodchild, M. Egenhofer, R. Fegeas and C. Kottman, Kluwer Academic**:** , 55-69.

Borner, K., C. M. Chen, et al. (2003). "Visualizing knowledge domains." Annual Review of Information Science and Technology **37**: 179-255.

Chen, C. M., J. Kuljis, et al. (2001). "Visualizing latent domain knowledge." IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews **31**(4): 518-529.

Chrisman, N. (1999). "A transformational approach to GIS operations." International Journal of Geographical Information Science **13**(7): 617-637.

Ding, Y. and S. Foo (2002). "Ontology research and development. Part I - a review of ontology generation." Journal of Information Science **28**(2): 123-136.

Ding, Y. and S. Foo (2002). "Ontology research and development. Part 2 - a review of ontology mapping and evolving." Journal of Information Science **28**(5): 375-388.

Fabrikant, S., Ed. (2000). The geography of semantic information spaces. GIScience 2000: The First International Conference on Geographic Information Science. Savannah, Georgia, USA, University of California Regents.

Fonseca, F., M. Egenhofer, et al. (2002). "Semantic granularity in ontology-driven geographic information systems." Annals of Mathematics and Artificial Intelligence **36**(1-2): 121-151.

Frank, A. (2001). "Tiers of ontology and consistency constraints in geographical information systems." International Journal of Geographical Information Systems **15**(7): 667-678.

Gahegan, M. (2003). "Is inductive machine learning just another wild goose (or might it lay the golden egg)?" International Journal of Geographical Information Science **17**(1): 69-92.

Guarino, N. (1995). "Formal ontology, conceptual analysis, and knowledge representation." International Journal of Human-Computer Studies **43**: 625-640.

Harvey, F., W. Kuhn, et al. (1999). "Semantic interoperability: A central issue for sharing geographic information." Annals of Regional Science **33**(2): 213-232.

Kavouras, M. and M. Kokla (2002). "A method for the formalization and integration of geographical categorizations." International Journal of Geographical Information Science **16**(5): 439-453.

Kavouras, M., M. Kokla, et al. (2005). "Comparing categories among geographic ontologies." Computers & Geosciences **31**(2): 145-154.

Kokla, M. and M. Kavouras (2001). "Fusion of top-level and geographical domain ontologies based on context formation and complementarity." International Journal of Geographical Information Science **15**(7): 679-687.

Kuhn, W. (2005). Geospatial semantics: Why, of what,and how? Journal on Data Semantics Iii. Berlin, SPRINGER-VERLAG BERLIN. **3534:** 1-24.

Lee, J. L., S. E. Madnick, et al. (1996). "Conceptualizing semantic interoperability: A perspective from the knowledge level." International Journal of Cooperative Information Systems **5**(4): 367-393.

Leung, Y. (1997). Intelligent Spatial Decision Support Systems. Berlin, Springer-Verlag.

Musen, M. A. (1998). "Modern architectures for intelligent systems: Reusable ontologies and problem-solving methods." Journal of the American Medical Informatics Association: 46-52.

Pike, W. and M. Gahegan (2003). Constructing semantically scalable cognitive spaces. Spatial Information Theory, Proceedings. Berlin, Springer-Verlag. **2825:** 332-348.

Pundt, H. and Y. Bishr (2002). "Domain Ontologies for data sharing - an example from environmental monitoring using field GIS." Computers & Geosciences **28**: 95-102.

Raper, J., J. Dykes, et al. (2002). "A framework for evaluating geographical information." Journal of Information Science **28**(1): 39-50.

Robinson, V. B. and D. S. Mackay (1995). "Semantic modeling for the integration of geographic information and regional hydroecological simulation management." Computers Environment and Urban Systems **19**(5-6): 321-339.

Rosch, E. (1978). Principles of categorization. Cognition and Categorization. E. Rosch and B. Lloyd. Hillsdale, N.J., Lawrence Erlbaum Associates**:** 27-77.

Sheth, A. P. and J. A. Larson (1990). "Federated database systems for managing distributed, heterogeneous, and autonomous databases." ACM Computing Surveys (CSUR) **22**(3): 183-236.

Skupin, A. (2004). "The world of geography: Visualizing a knowledge domain with cartographic means." Proceedings of the National Academy of Sciences of the United States of America **101**: 5274-5278.

Stock, K. and D. Pullar (1999). Identifying semantically similar elements in heterogeneous spatial databases using predicate logic expressions. Interoperating Geographic Information Systems. Berlin, SPRINGER-VERLAG BERLIN. **1580:** 231-252.

Tversky, B. and K. Hemenway (1984). "Objects, parts, and categories." Journal of Experimental Psychology **113**(2): 169-193.

Uschold, M. (1998). "Knowledge level modelling: concepts and terminology." Knowledge Engineering Review **13**(1): 5-29.

Viger, R. J. (2004). GEOLEM: Improving the Integration of Geographic Information in Environmental Modeling Through Semantic Interoperability. Geography. Boulder, University of Colorado**:** 135.

Visser, U., H. Stuckenschmidt, et al. (2002). "Ontologies for geographic information processing." Computers & Geosciences **28**(1): 103-117.

Voisard, A. and H. Schweppe (1998). "Abstraction and decomposition in interoperable GIS." International Journal of Geographical Information Science **12**(4): 315-333.

Worboys, M. F. and E. Clementini (2001). "Integration of imperfect spatial information." Journal of Visual Languages and Computing **12**(1): 61-80.

Project Support

Cooperators/Collaborators

- Dr. Barbara Buttenfield, Professor, Dept. of Geography, University of Colorado – Boulder, Guggenheim 110; 260 UCB, Boulder, CO 80309-0260, (303)492-3618, babs@colorado.edu
- Dr. Olaf David, Computer Scientist – Colorado State University/USDA Agricultural Research Service, 2150 Centre Avenue, Building D, Suite 200, Fort Collins, CO 80526-8119, (970) 492-7312, olaf.david@ars.usda.gov
- Frank Geter, ITC Lead Modeling Specialist, Natural Resources Conservation Service, 2150 Centre Avenue, Building D, Suite 200, Fort Collins, CO 80526-8119, (970) 492-7370, fgeter@ftc.usda.gov
- Jeff Hamerlinck, Director of the Wyoming Geographic Information Science Center, Research Scientist, Department of Geography, University of Wyoming, Agriculture C, Room 306, University of Wyoming, Laramie, WY, (307) 766-2736, Itasca@uwyo.edu
- Dr. Charles G. O'Hara, Research Professor, GeoResources Institute, Mississippi State University, 2 Research Blvd., ERC, Room 201, Starkville, MS, 39762, (662) 325-2067, cgohara@gri.msstate.edu
- Roland Viger, Physical Scientist – USGS, MS 412 Box 25046 DFC, Denver CO, 80225, (303) 236-5030, rviger@usgs.gov
- The U.S. Federal Interagency Steering Committee on Multi-Media Environmental Modeling (www.iscmem.org). Note that GEOLEM is already associated with this group, at http://iscmem.org/WG01_Geolem.htm.

Other Project Support

- Roland Viger completed (2004) a Masters thesis establishing the fundamental GEOLEM concepts and is currently extending these for his PhD dissertation. He hopes to complete the PhD program within a year.
- Dr. David, working with the USDA ARS, has committed to provide an operational installation of the GEOLEM for the NRCS National Weather and Climate Center in support of the their water supply forecasting duties. To this end, Dr. David has already hired a contractor to work on the GEOLEM development effort full-time for the next five months. A portion of funds from the CEGIS project will be used to extend the contractor's funding. GEOLEM is expected to serve as a geo-services infrastructure for the USDA Object Modeling System, which is a nation-wide system for development and deployment of simulation modeling capacity to field workers.
- Dr. O'Hara will utilize the GEOLEM approach as part of his research on behalf of NASA efforts to rapidly deploy new satellite data streams into scientific computing applications, particularly simulation modeling. A Joint Funding Agreement has been established between the USGS and Dr. O'Hara to support Roland Viger's continued development of GEOLEM at a funding level of $150,000 over the next 18 months.

- The EPA provided (2003) $90,000 of seed-money for the initial research into GEOLEM, coordinated through Roland Viger, Dr. David, and Dr. Luis Garcia of Colorado State University.

Budget

This is a one-year project.

Budget Request

Fiscal Year 2007 Budget						Total Year 1
	4556	U Colo-Boulder w/ student	MS State	Colo State U w/ student	U of Wyo	
Personnel Salary	34,000	21,089	12,000	23,339	12,000	102,428
Other expenses	2,500	2,000	2,400	3,100	1,000	11,00
TOTAL DIRECT	36,500	23,089	14,400	26,439	13,000	**113,428**
Gross Assessment Rate	16%	45.52% (estimated)	43.5%	22.69%	50% (estimated)	
INDIRECT COSTS ESTIMATE	5,840	5,167	6,264	12,417	6,500	36,188
TOTAL	42,340	28.256	20,664	38,856	19,500	**149,616**