Geoinformatics 2006—Abstracts

May 10–12
Reston, Virginia

Scientific Investigations Report 2006-5201

U.S. Department of the Interior
U.S. Geological Survey

Geoinformatics 2006—Abstracts

Edited by Shailaja R. Brady, A. Krishna Sinha, and Linda C. Gundersen

Scientific Investigations Report 2006-5201

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Dirk Kempthorne, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2006

For sale by U.S. Geological Survey, Information Services
Box 25286, Denver Federal Center
Denver, CO 80225

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: http://www.usgs.gov/

Suggested citation:
Brady, S.R., Sinha, A.K., and Gundersen, L.C., editors, 2006, Geoinformatics 2006—Abstracts: U.S. Geological Survey Scientific Investigations Report 2006-5201, 60 p.

Preface

In order to facilitate the discovery, integration, and analysis of distributed data, geoscientists—in partnership with information technologists and computer scientists—have established the emerging science of geoinformatics. Geoinformatics 2006, an international conference that was attended by over 300 participants, represents the first nationally organized meeting whose primary goal was to provide a forum for the exchange of the most current research in geoinformatics.

It is a pleasure to acknowledge the sponsorship of the U.S. Geological Survey (USGS), National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), Geological Society of America (GSA), and the Geosciences Network (GEON). On behalf of all the participants of Geoinformatics 2006, we appreciate the support of the U.S. Geological Survey headquarters staff in Reston, Va., for providing facilities and technical support for the conference, of GEON staff for managing the registration process and the abstracts, and staff from the San Diego Supercomputer Center for videotaping the conference.

Organizing Committee:
Margaret Banton (Chair), San Diego Supercomputer Center, University of California—San Diego; Linda C. Gundersen, U.S. Geological Survey; and A. Krishna Sinha, Virginia Polytechnic and State University

Technical Committee:
Randy Keller (Chair), University of Oklahoma—Tulsa; Robert Raskin, Jet Propulsion Laboratory; and Chaitan Baru, San Diego Supercomputer Center, University of California—San Diego

Technical support by U.S. Geological Survey:
Judith Back, Jenifer Bracewell, Esmeralda Campos, Kimberly Dueling, Donna Foulke, Martha Kiger, Susan Kropschot, Troy Miller, Cynthia Myers, Carol Quesenberry, Vitmary Rodriguez, Edward Simmons, Elizabeth Stettner, June Thormodsgard, and Diane Welch.

Technical support by GEON (Geosciences Network):
Margaret Banton and Nancy White

Webcasting services by San Diego Supercomputer Center (SDSC):
Sebastian Allen and Emilio Valente

Financial support was provided by the U.S. Geological Survey, the National Science Foundation's Division of Earth Science (EAR), Environmental Systems Research Institute (ESRI), San Diego Supercomputer Center, and the California Institute for Telecommunications and Information Technology

For the complete program, Webcasts of the keynote speeches and technical program, slides, and poster presentations, please go to http://www.geongrid.org/geoinformatics2006/.

Contents

Geoinformatics 2006—Abstracts

Edited by Shailaja R. Brady, A. Krishna Sinha, and Linda C. Gundersen

Session 1—Ontologies and the Semantic Web

Chair: Robert Raskin, Jet Propulsion Laboratory, Pasadena, Calif.

Solar-Terrestrial Ontologies

By Deborah L. McGuinness[1], Peter A. Fox[2], Luca Cinquini[3], and Don Middleton[3]

[1]Knowledge Systems, Artificial Intelligence Laboratory/Stanford University, and McGuinness Associates, Stanford, Calif.

[2]High Altitude Observatory/Earth and Sun Systems Laboratory/National Center for Atmospheric Research, Boulder, Colo.

[3]Scientific Computing Division, Computational and Information Systems Laboratory/National Center for Atmospheric Research, Boulder, Colo.

Ontologies are being used to enable environments that support scientific data sharing and interoperability. Virtual observatories, such as the Virtual Solar Terrestrial Observatory (VSTO; home page at http://vsto.hao.ucar.edu/), are using ontologies to support search and retrieval services that can use the meaning of terms in the search queries instead of just the syntax of the terms. In the VSTO project, we have implemented a prototype portal to access two very different datasets: one from solar physics (images of the solar atmosphere; see the Mauna Loa Solar Observatory home page at http://mlso.hao.ucar.edu/) and another from terrestrial upper atmospheric physics (time series of thermodynamic and dynamic parameters; see the Coupling, Energetics, and Dynamics of Atmospheric Regions (CEDAR) home page at http://cedarweb.hao.ucar.edu/).

This portal is implemented within a semantic framework utilizing leading-edge semantic Web technologies and methodologies. A cornerstone of this operational data framework is an ontological representation of the solar-terrestrial concepts, such as instruments, parameters, date and time, and datasets. We present the current state of our efforts, the methods, tools and technologies utilized as well as near-future steps.

Cyberinfrastructure for the Geosciences—Ontology-Based Discovery and Integration

By A. Krishna Sinha[1], Kai Lin[2], Robert Raskin[3], and Calvin Barnes[4]

[1]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Va.

[2]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[3]Jet Propulsion Laboratory, Pasadena, Calif.

[4]Department of Geosciences, Texas Tech University, Lubbock, Tex.

One of the significant challenges towards a more comprehensive understanding of processes and events that have shaped the earth though time is the need to bring disparate datasets together through a common Web-based smart query system. Data discovery is clearly a prerequisite for integration, and we report on new techniques that support discovery and integration of geoscience data. We propose a three-tier database registration system that facilitates discovery of data, as well as integration across subdisciplinary databases. At the first level we utilize a controlled vocabulary based on American Geological Institute (AGI) index terms with plans for cross indexing to vocabularies in the National Aeronautics and Space Administration's (NASA's) Global Change Master Directory (GCMD) and the U.S. Geological Survey's North American Datum Model (NADM). All known geoscience databases could be registered (through the Geosciences Network (GEON) portal at www.geongrid.org) using these terms for rapid discovery, and used by research scientists, educators, and students. The second- and third-tier registration of databases is based on application of domain ontologies. We have developed new ontology-based packages containing Planetary Materials (elements, isotopes, rocks, and minerals), Planetary Structure, Planetary Location, Planetary Phenomenon, Physical Properties, Geologic Time, and Geo Images. Through application of ontologies, we utilize smart-search techniques to find datasets within large data resources. For example, it is possible to identify rock types in datasets even though they are not included in the metadata, and thus extract the related datasets for further analysis.

Integration of databases can be done through (1) schema merging, when the user is knowledgeable about data organization (semantics of the schema), (2) view-based techniques that include the creation of a virtual schema to allow the user to address structural heterogeneity, or (3) ontology-based integration, accomplished by registering databases to ontologies. We favor ontology-based integration because it systematically resolves both syntactic and semantic heterogeneity, allowing scientists to focus on the content of the database rather than its schema. It is now possible to register distributed databases to data ontologies, making possible integration of multiple distributed databases into a single virtual database. An added benefit of this approach is that workflow-based computations can interact with multiple databases.

Designing Dynamic Earth Sciences Ontologies

By Hassan A. Babaie[1]

[1]Department of Geosciences, Georgia State University, Atlanta, Ga.

Depiction of complex spatio-temporal geological processes (for example, frictional sliding and crystallization) and objects, such as fault, cataclasite, and mineral that participate in them, requires developing ontologies that realize both perdurant (occurrent) and endurant (continuant) entities that constitute most phenomena in the real world. Endurant entities include spatial things such as magma and vein that persist identically, in full (that is, with all mereological, spatial parts) in every instant of their existence, and through change (for example, the Andes mountains retain their identity over their life span.) Ontologies depict endurants at some specific time along the time continuum. Perdurant entities, such as recrystallization and melting, on the other hand, lack spatial parts, and instead unfold themselves over time in successive phases that are considered as temporal parts. Process perdurants include material and immaterial entities as participants and agents. This means that they only exist in successive phases; their past phases don't exist any more. For example, the earlier phase of magma mixing does not exist during the later phase of magma crystallization. Stages constitute the instantaneous parts of perdurants, which also include the time slices of the lives of endurants. The life of an endurant is a perdurant; that is, it unfolds over a period between t_1 and t_n (for example, the time interval in which a fault is formed, sticks, slides, and is dead, in different regions). The two types of spatio-temporal entities are located in spacetime, compared with abstract entities that do not occur in spacetime.

Domain ontologies traditionally take one of two perspectives of the world, either depicting successive instantaneous snapshots of entities (SNAP ontologies), or viewing processes by spanning an interval of time (SPAN ontologies) (Grenon and Smith, 2004). In the first perspective, based on three dimensionalism, change is measured by comparing discrepancies in the snapshots at different times. The SPAN perspective, based on four dimensionalism, is more realistically capable of depicting events, processes, and spatio-temporal regions. The two perspectives are indeed complementary and must be used in combination in the depiction of complex domains in earth science. Granularity, the scale at which we view the world, is another dimension which is orthogonal to the two perspectives, allowing the SNAP/SPAN ontologies to depict the world at different levels (for example, microscopic fault to lithospheric fault).

The regions in which the endurants exist are part of the four-dimensional space that can have any dimension (<5; including fractal), shape, and size. Endurants undergo changes in their mereological (that is, parthood) structure; for example, a strike-slip fault develops new bends, sag ponds, and step parts. Endurants may also have variable topology over their lifetime; for example, faulting brings two particular rock bodies in contact with, or above, each other, thereby changing their connection, adjacency, coincidence, and overlap relations. Infiltrating water may bring a contaminant entity into the immaterial voids (pores) of rock, modifying the topology of the rock water interface (water co-locates with the pore).

Mereologically, endurants have other endurants as parts, but perdurants have other perdurants and stages as part. For example, brittle deformation includes the sub-process of cataclasis, which includes comminution, rotation, and frictional sliding at different time intervals. Each of these processes has its own multiple stages. Ontologies can model the mereotopological relationships between real entities through formal relations such as parthood, connection, dependence, identity, and difference.

In this paper, I present the two perspectives of ontologies on the real world, and discuss formalization of process ontologies, applying mereotopology. Indeed, the very definition of geologic structure that portrays the orientation, spatial position and distribution of discontinuity parts of rocks and regions (at all levels of granularity, from crystal lattice to lithosphere scale), borrows from mereology and topology. Earth science ontologies can benefit immensely from formal relations based on mereotopology.

Reference Cited

Grenon, P., and Smith, B., 2004, SNAP and SPAN—Towards a dynamic spatial ontology: Spatial Cognition and Computation, v. 4, no. 1, p. 69-103.

The Taxonomic Synonymy Definition Framework—An Implementation of the Resource Description Framework for the Exchange of Taxonomic Synonymy

By Douglas Fils[1], Iain Matcham[2], Cinzia Cervato[1], and James Crampton[2]

[1]CHRONOS, Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

[2]Institute of Geological and Nuclear Science Limited, Lower Hutt, New Zealand

CHRONOS, in collaboration with Geological and Nuclear Sciences New Zealand (GNS), has begun development and implementation of a taxonomic synonymy definition framework (TSDF). Based on the resource description format (RDF), the TSDF defines a mechanism to codify and exchange synonymy concepts in RDF using principles and concepts from the simple knowledge organization system (SKOS).

TSDF utilizes predefined community namespaces like Dublin Core, Friend of a Friend (FOAF) and others as much as possible in addition to SKOS. Careful use of namespaces allows for the gleaning of data, extensibility by others, and potential interaction with other synonymy encoding schema.

The goal is to provide a navigable interface to synonymy data for the researcher as well as a standard means to exchange synonymy concepts between institutions. Implementation in RDF allows for the use of faceted interfaces to RDF data such as Longwell from the Massachusetts Institute of Technology (MIT) Simile Project as well as the rich set of RDF tools from the community. These interfaces allow for easy navigation of the various relationship paths between concepts in an RDF document. Practices from the Semantic Web Environmental Directory project, part of the Semantic Web Advanced Development (SWADEurope), are being used to guide the development of an architecture for a distributed community of synonymy providers. The goal is the development of a community of synonymy providers able to publish and harvest concepts from each other for use in searches or other taxonomy-related research.

TSDF is a method to represent synonymy data. It is not the scope of TSDF to define a method to store or query data when the use of Web services and representative state transfer (ReST) architectures can quite easily address the need for requesting and transporting data in TSDF format. A query of TSDF -formatted files can be done in many ways with SPARQL (recursively, SPARQL Protocol and RDF Query Language), which is one of the most common methods to extract data from RDF graphs. This exchange of data is facilitated without imposing obligations on data storage or network application structures of the host institutions. Applications of unique identifiers along with predefined namespaces are used for the identification of taxonomic concepts within TSDF. The use of preexisting practices provides an environment where TSDF data can be widely and easily accessed by a large community of users and providers. Details on the use of these existing practices will be given.

The presentation will cover (1) the current state of TSDF implementation, giving examples of its use to exchange data between CHRONOS and GNS, and (2) the initial implementation of interfaces to the data collections. GNS has led the development of a set of Java-based tools and application program interfaces (APIs) related to TSDF and these will be outlined. CHRONOS is implementing an architecture and interface to available data using these resources that will also be detailed. An architecture using Web services and ReST to arbitrate requests and transportation of TSDF-encoded synonymy data is detailed in the presentation. A discussion of the application of existing architectures for the development of a distributed synonymy network will be included.

Using Semantic Web Modeling and Technology for Automated Search for Data

By Nancy Wiegand[1]

[1]University of Wisconsin—Madison, Madison, Wis.

Searching for data sources over the Web, including using portals that record metadata, is problematic. User-entered keywords often do not return needed data sources, or return many irrelevant results. Instead, data specialists may resort to general Internet searches or a sequence of phone calls. Better methods are needed to retrieve data, and, ideally, the process would be automated. We explore the abilities and limitations of Semantic Web languages, tools, and rule engines for this purpose.

We observe that many searches are for tasks that need the same types of data with just the location varying. For example, tasks involving emergency response or land-use planning need similar data but for different places. To solve the problem of finding data sources and to automate the process, we create a model to formalize a knowledge base and associations within it. Our model is task based and includes developing an ontology of tasks, derived from the symbology of incidents developed by the Homeland Security Working Group, and an ontology of data sources derived from the Federal Geographic Data Committee's (FGDC's) framework data. We also import the FGDC Web Ontology Language (OWL) metadata ontology and include a place taxonomy. Correspondences between types of data sources needed for types of tasks are established using examples from emerging reports of emergency task forces. This information is then formally recorded using OWL within Protégé, an ontology editor. Instance data is added for evaluation.

Although we attempted to fully design our model using OWL restrictions to keep all information within an established knowledge base, we found that geospatial data have too many selection criteria, such as scale, date, and accuracy, for that to be feasible. That is, full modeling by declaring numerous restrictions using OWL's "someValuesFrom" construct within anonymous subclasses was very cumbersome and inflexible. The base relationships between types of tasks and data sources, however, were still modeled in this manner to avoid hard-coding criteria in rules or queries as much as possible. Also, for simplicity, we expressed base relationships using a "needs" object property rather than creating an overarching class as done in various other Protégé applications. As an example, a "someValuesFrom restriction" for a fire emergency is "∀ needs (Road U LandCover U Hydrography)."

To specify further selection criteria, we needed to augment our system with Jess, a rule engine, which is accessible in Protégé through the JessTab plugin. Jess rules allow any number of selections and the use of comparison operators. However, due to the lack of JessTab's ability to translate OWL object properties to Jess assertions, we had to create general instance data for class types and re-set restriction information using instances. This then did allow the inferencing we needed.

In summary, our user interface allows selection of a task from the task ontology and a location from a place ontology. Through inferencing over base relationships modeled as OWL restrictions, the types of data sources needed for the selected task are returned. Using further rules that are pre-established, additional criteria are specified for each data source type, allowing the return of Web addresses to needed data sources.

The advantage of using ontologies rather than a database query approach is that base information can be re-used for other applications. The advantage of a formalized and automated system is that, in an emergency for example, anyone can quickly retrieve needed data sources.

Further, a full expert system could execute subsequent rules if particular criteria from initial rules do not yield a data source (for example, to find a later year or smaller scale). Such a system can be employed within existing portals or can stand alone on the Web using emerging distributed technologies including the OWL Query Language (OWL-QL).

GEONSearch—From Searching to Recommending

By Ullas Nambiar[1], Bertram Ludaescher[1], Ghulam Memon[2], and Dogan Seber[2]

[1]Department of Computer Science, University of California—Davis, Davis, Calif.

[2]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

The Geosciences Network (GEON) project is a large-scale collaborative effort aimed at creating a cyberinfrastructure for the earth sciences. GEON is a repository of heterogeneous data that can be integrated, queried, analyzed, and visualized by geoscientists with GEON-provided ontology-enabled tools. A critical first step in efficient retrieval and integration of data from GEON is the ability to identify the datasets of interest. GEONSearch—a component of the GEON Portal, plays a very important role in GEON by enabling users to discover datasets of interest by searching for a "Title" and "Description" of the dataset, spatial and temporal metadata, and mapped concepts from GEON ontology. While GEONSearch does enable searching over many associated features, the underlying search model is keyword based. Specifically, GEONSearch currently returns only datasets whose features explicitly match the keywords provided by the user. Thus, effective retrieval of datasets from GEONSearch will require users to construct keyword queries that clearly identify their need—a task that is often difficult even for experts. By focusing on exact keyword matches, GEONSearch would miss providing datasets that are relevant to the user query but whose metadata do not match the query. For example, if users search for "California," they will not find relevant documents that mention only "San Diego."

Similar problems could arise when searching over temporal metadata. A concept-level search provided by GEONSearch could help overcome this problem to a certain degree; however, the effective use of concept-based search will require the user to formulate the query in terms of available concepts. Hence, our focus is in improving the relevance of results provided by the easy-to-use keyword search currently provided by GEONSearch. We will do so by incorporating background knowledge available from GEON ontologies, dataset-to-ontology mappings, past usage information, and the content (schema and data) overlap between sources.

The primary motivation behind this talk is to avoid the "frog in a well" syndrome of doing research where the tools developed do not meet user needs. We provide an overview of the research being done to improve GEONSearch with the intent of receiving feedback from the users (geoscientists). In particular, we are interested in obtaining "use cases" that will be helpful to our research and (or) pose additional challenges to be met.

In this talk, we will present results of ongoing research that will allow GEONSearch to learn the relevance of a dataset to a query, to other datasets, and to GEON ontologies. With this knowledge GEONSearch will be able to recommend related datasets and ontologies to users once they pick a dataset of interest. More importantly, GEONSearch would be able to provide justifications for its recommendations that are based on user-accepted semantics. This ability to learn related or relevant datasets would also be useful in improving the data registration process by allowing GEON to suggest and (or) automatically map datasets to those relevant ontologies that the user may have missed.

Knowledge Sifter—Agent-Based Search Over Heterogeneous Sources Using Semantic Web Services

By Larry Kerschberg[1], Daniel A. Menascé[1], and Hanjo Jeong[1]

[1]E-Center for E-Business, George Mason University, Fairfax, Va.

Knowledge Sifter is a scaleable agent-based system that supports access to heterogeneous information resources such as the Web, open-source repositories, Extensible Markup Language (XML) databases and the emerging semantic Web. The Knowledge Sifter agent architecture has three layers: user layer, knowledge management layer, and data layer. The user agent interacts with the user to elicit user preferences that are managed by the preferences agent. These preferences include the relative importance attributed to terms used to pose queries, the perceived authoritativeness of Web search engine results, and other preferences to be used by the ranking agent.

The user indicates an initial query to the query formulation agent. This agent, in turn, consults the ontology agent to refine or generalize the query based on the semantic mediation provided by the ontology services available. The ontology agent uses a conceptual model for the domain by means of the Web ontology language (OWL) schema specification of the Imagery Domain Model (based on ISO 19115 and 19139 of the International Standards Organization).

The Knowledge Sifter architecture is general and modular so that new ontologies and new information resources can be easily incorporated. The Web services agent uses domain knowledge regarding the data sources (such as quality of service (QoS) attributes, source authoritativeness, and image sizes) to optimize the execution of subqueries. The ranking agent is responsible for compiling the subquery results from the various sources, ranking them according to user preferences supplied by the preferences agent.

The Knowledge Sifter project is sponsored by a National Geospatial-Intelligence Agency (NGA) University Research Initiatives (NURI) through their InnoVision Program. The NGA Web page may be found online at http://www.nga.mil/portal/site/nga01/index.jsp?front_door=true/. Drs. Kerschberg and Menascé are co-principal investigators on this project. Mr. Hanjo Jeong is the Chief Programmer for Knowledge Sifter. For more information about Knowledge Sifter, please visit George Mason University's E-Center for E-Business publications page online at http://eceb.gmu.edu/publications.html/.

Session 2—Geologic Mapping and Databases

Chair: Peter T. Lyttle, U.S. Geological Survey, Reston, Va.

Spatial Data Infrastructure—What Is It? Why Is It Critical?

By Alan R. Stevens[1]

[1]Federal Geographic Data Committee, U.S. Geological Survey, Reston, Va.

The confidence that citizens have in their governance and its ability to make reliable everyday decisions, as well as efforts to avert future tragedies, might improve by understanding the role that geographic information and related spatial technologies play in bringing order to the chaos. Moreover, these technologies relate to the application of geographic data and information into the day-to-day decisionmaking, as well as directly assisting government and industry planning for homeland security, critical infrastructure assurance, and other public health and safety challenges. As such, knowledge about the application of geographic information systems (GISs) has become prevalent. The orderly application of these data and techniques requires the investment in a spatial data infrastructure (SDI) at the local to national levels and partnerships among the governmental units, nongovernmental organizations (NGOs), academia, and the private sector. An SDI provides a common reference system, common standards, and a common language.

The SDI message is simple and over arching: GISs will not work unless the data fits together horizontally and vertically, and there are more chances that it will not fit together under most conditions today. As an example, if one studies the road center-line files from Federal, State, county, city, utility and other sources, the chances of them not matching ground truth by 1, 5, 10, 30, 50, 100 meters or more are very realistic. Furthermore, there will also be a chance that the following types of data will not match either, nor will they register to the road center-line file: property boundaries; sewer, water, electric, and gas lines; subway tunnels; building footprints; zoning boundaries; watersheds, land use; water, gas, and oil wells; land fills, and related phenomena. If data do not fit together, the user must either convert or even re-collect the same data at great additional time and expense; time and money are often in very short supply, particularly in emergency situations.

The solution to this problem is to encourage organizations that collect, process, archive, distribute, and use geographic data and information to do so using common standards and interoperable systems and processes. "Common data standards" means that one can easily convert data files from one source to fit with data from other sources. "Interoperable systems and processes" means that users can collect data using one hardware system, and the resultant data can be used with another provider's software. Typically the data obtained from the Federal Geographic Data Committee (FGDC)-registered clearinghouses have met accepted standards and have been collected by interoperable systems and techniques. Some do not, but the information is available in the metadata records without going through time consuming trials of analyzing full datasets. Clearinghouses from many organizations throughout the world can be found at the FGDC Clearinghouse Registry online at http://registry.gsdi.org/.

Advances in Design of the U.S. National Geologic Map Database

By David R. Soller[1], Stephen M. Richard[2], Jon Craigue[3], and David Percy[4]

[1]U.S. Geological Survey, Reston, Va.

[2]Arizona Geological Survey, Tucson, Ariz.

[3]University of Arizona and U.S. Geological Survey, Tucson, Ariz.

[4]Portland State University, Portland, Oreg.

The U.S. Geological Survey and the Association of American State Geologists (AASG) are mandated by Congress to provide a National Geologic Map Database (NGMDB, http://ngmdb.usgs.gov/) of standardized geoscience information that can be used to address societal issues and to support scientific research. This partnership serves to advance (1) the goal of building the NGMDB, and (2) the need for each geological survey to improve their ability to deliver map information to their users.

Because of the mandate's broad scope, a phased, incremental approach is used to design the NGMDB. This approach allows for the necessary time to build consensus and expertise among the State geological surveys and the USGS. Furthermore, it helps the database designers to more effectively consider and respond to evolving technology and user needs. Since 1996, NGMDB designers have (1) created a Geoscience Map Catalog and Image Library, (2) extensively revised the U.S. Geologic Names Lexicon and made it Web-accessible, and (3) provided a database of current mapping projects, all in order to quickly provide users with basic geoscience information.

The ultimate goal for the NGMDB project is to design and build an online database that contains a national archive of richly attributed, standardized geoscience information in vector and raster formats. That database is intended to be a distributed system, with nodes hosted by the State geological surveys and the USGS, and integrated with the existing NGMDB databases described above. The standards necessary to build this system are under development, in cooperation with other U.S. and Canadian agencies; our principal partners include the Geological Survey of Canada, the North American Data Model (NADM) Steering Committee, the International Union of Geological Sciences (IUGS) Commission for the Management and Application of Geoscience Information (CGI), and the U.S. Federal Geographic Data Committee (FGDC). These collaboratively developed draft standards include the following: a science terminology (developed by NADM's Science Language Technical Team); a conceptual data model (NADM Conceptual Model 1.0); an implementation of the data model for the NGMDB that is compatible with Environmental Systems Research Institute's (ESRI's) Geodatabase; a map data interchange schema written in GeoSciML (based on Geographic Markup Language (GML) and supported by the CGI); a FGDC-supported terminology for describing locational accuracy of geologic features in the field; and FGDC-supported symbols, patterns, and colors for geologic maps. These standards and technical guidance can be found online at http://ngmdb.usgs.gov/Info/ (see especially the "Digital Mapping Techniques Workshop Proceedings").

Current research and development includes a data-entry and database-management software tool with an accompanying suite of standard science terminologies, the prototype database implementation in ESRI's Geodatabase format, and participation in the CGI's GeoSciML development, including the Interoperability Testbed, which is described on the CGI Web page online at https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/TestBed2/.

GeoSciML—An International GML 3 Standard for Sharing Geologic Map Information, With Examples from Canada and the U.S.

By Boyan Brodaric[1], Eric Boisvert[2], Bruce R. Johnson[3], and Simon Cox[4]

[1]Geological Survey of Canada, Ottawa, Canada

[2]Geological Survey of Canada, Quebec, Canada

[3]U.S. Geological Survey, Reston, Va.

[4]Commonwealth Scientific and Industrial Research Organisation, Perth, Australia

Open geospatial Web services are the foundation for national spatial data infrastructures in Canada, the U.S., and in several other countries. Many geological data providers are experimenting with these services in relative isolation, leading to a wide disparity in the format of the served data. The International Union of Geological Sciences' Commission for the Management and Application of Geoscience Information (IUGS-CGI) Data Model Collaboration (DMC) is a significant international collaboration intended to harmonize the various Web services efforts of national geologic data providers. In the past two years the DMC has (1) developed GeoSciML, a sophisticated GML 3 application for geology, and (2) initiated two multi-country testbeds to evaluate it and related Web service technologies. In this paper we describe the development of GeoSciML, including its derivation from unified modeling language (UML) conceptual models, and report on the second international testbed involving the provision of geologic map information by six countries. We focus on an implementation from the Geological Survey of Canada and the U.S. Geological Survey, showing how a geologic map from each of these organizations can be served as GeoSciML using open geospatial Web services. We also discuss challenges in transforming geologic map databases to GeoSciML. We are optimistic that GeoSciML is poised to be a major delivery vehicle for geoscience information.

Dynamic Digital Maps—An Open-Source Tool to Distribute Maps, Data, Articles, and Multimedia as an Integrated Stand-Alone Cross-Platform Package Via the Web and CD/DVD for Use in Research, Teaching, and Archiving Information

By Christopher D. Condit[1]

[1]Department of Geosciences, University of Massachusetts, Amherst, Mass.

Dynamic digital maps (DDMs) are computer programs that distribute high-quality color maps, digital images, movies, analytical data, and explanatory text, including collar text and field guides, in integrated cross-platform, Web-enabled formats that are intuitive to use, easily and quickly searchable, and require no additional proprietary software to operate. DDMs provide an ideal format in which to present research, archive data, and create a rich teaching resource. The format includes a home page containing a graphic index map, from which detailed maps can be opened, and buttons to palettes containing the DDM's indexes, each of which also links to the material they list.

An open-source program, the "DDM-Template" (into which you can insert your data), and an accompanying "cookbook" on how to do this are available online at the University of Massachusetts' Department of Geosciences Web page, http://ddm.geo.umass.edu/, along with numerous DDMs that demonstrate this potential. Making a DDM from the template requires the use of the relatively inexpensive, multiplatform programming environment Runtime Revolution (available on the Web at http://www.runrev.com), which has a low learning curve. Maps and photos (saved as JPEG files) and movies (including QuickTime panoramas) are stored outside the program, which acts as an organizational framework and index to present these data. Text files (for example, map collar text, field trip guides, descriptions of geologic settings, and image captions) are saved within the program, and can be included in three different versions, which the user selects when setting preferences. A single DDM can thus be aimed at audiences with three differing levels of scientific training, or in three different languages. Analytical data are uploaded and stored as tab-delimited text within the program and can be saved as text documents for use out of the program, or for inclusion in traditional databases. Map-unit labels, sample sites, and graphics, such as camera icons (created in the template), can be overlaid on top of maps or images; a link can be provided to view the associated data or images. Once data have been added to the "DDM-Template," or stored in specified directories, a single step creates royalty-free stand-alone programs for numerous Unix, and all Windows and Macintosh operating systems. The DDM program matching the user's operating system can be downloaded from a Web page. The DDM program can then access its associated data directly from that site with no browser needed. Alternatively, the entire package can be distributed and used from CD, DVD, or from flash-memory storage. The open-source DDM-Template could be modified to access files stored in databases and to open maps from XML files.

DDMs can provide an essential context for teaching and interpreting the local geology and for presenting a framework in which to imbed field trips for classes or professional meetings; examples can be downloaded from the University of Massachusetts Department of Geosciences' Web page at http://ddm.geo.umass.edu. The DDM of New England, which contains six field trips, and the Massachusetts State Geologic Map typify such a use. DDMs also can provide a framework for interpreting geochemical, geophysical, and geochronological information; an example of such use can be found in the DDM of the Springerville volcanic field. Using this DDM, one can examine the analytical data of each sample within the context of its location on any map or image (including photomicrographs). This DDM also demonstrates the capability to search for all data for a given unit or sample, in addition to locating all occurrences of a unit or sample on the maps or images in the DDM. It also provided a framework for junior-level petrology exercises as described by Boundy and Condit (2004).

The Office of the Massachusetts State Geologist faces the task of making maps easily accessible to citizens in the most efficient and expeditious manner possible. Many of these potential users have neither the expertise nor software to use a geographic information system (GIS) product, but do have the basic computer skills to use an integrated stand-alone product such as the DDM. To accommodate that need, the Office is experimenting with the release of a preliminary version of the Marlborough quadrangle of Massachusetts as a DDM. Other more exotic examples include a DDM of the Moon, and one of Mars.

Reference Cited

Boundy, T.M., and Condit, 2004, Bringing the field into the classroom by using dynamic digital maps to engage undergraduate students in petrology research: Journal of Geoscience Education, v. 52, no. 4, p. 313-319.

Building a Dynamic Image-Based Database—Integrating Thin-Section Images and Data Using Dynamic Digital Maps

By Michael L. Williams[1], M. Cleveland[1], and Christopher D. Condit[1]

[1]Department of Geosciences, University of Massachusetts, Amherst, Mass.

The analysis and reanalysis of metamorphic tectonites and many other geological materials require the close linkage of optical and compositional images, microstructural observations and measurements, compositional analyses, and petrological models. The maintenance of this linkage, especially between images and data, is critical if other workers are to use existing data to answer new questions. We present an image-based data analysis and archival system with full integration of images, observations, and quantitative data. Dynamic digital maps (DDMs) have been developed to integrate geologic maps, field observations, and analytical data. We use full-thin-section images as DDM base maps for dynamic thin section databases. The images, in a variety of types and at different scales, are directly linked to datasets ranging from mineral analyses, whole-rock analyses, geochronologic information, and texture or fabric data. In addition, the DDMs are linked to tools for analysis of raw data and for modeling and integrating results. The linked tool set is a critical, value-added component, which makes the DDM both an archiving tool and also a powerful tool for analyzing and integrating data during the primary data collection and analysis stage of a project. Essentially, single thin sections, and their resulting pressure-temperature-time-deformation (P-T-t-D) histories become dynamic, multifaceted nodes in a regional tectonic database. Further, because the DDMs access raw analytical data (from microprobe traverses or geochronological analyses), new workers can re-analyze a specimen and resample data in new ways. This type of integration of multiscale images, observations, and raw data will be critical if tectonic data are to be archived and reused in the future. Using “DDM-Template,” DDMs are easy to create and modify without extensive geographic information systems (GIS) background. As an example, a DDM thin-section database has been generated from a single sample of folded Paleoproterozoic turbidite from the Upper Granite Gorge of the Grand Canyon in the Western United States. The DDM integrates structural, microstructural, petrologic, thermodynamic, and geochronologic information, all of which are critical for characterizing the P-T-t-D history. High-resolution images (including optical images, microprobe compositional maps, QuickTime movies, electron images, and interpretive sketches), ranging in scale from full-section to single mineral have been used to document texture, fabric, and compositional relationships. The integration of images, interpretations, and raw compositional data allows workers to analyze and archive data and interpretations from new or ongoing studies and will allow future workers to use existing data for constraining entirely new research questions and hypotheses.

Spatial and Temporal Pattern of Late Pleistocene and Holocene Fault Slip-Rates in the Alvord Extensional Basin, Southeastern Oregon

By John S. Oldow[1] and Eron S. Singleton[1]

[1]Department of Geological Sciences, University of Idaho, Moscow, Idaho

The Alvord extensional basin of southeastern Oregon provides an ideal setting to compare rates of contemporary displacement to the spatial and temporal pattern of intermediate-time-scale fault slip. The Alvord basin lies near the northern extent of the belt of active deformation that is localized along the western margin of the Great Basin. Active deformation in the Alvord basin is reflected in the physiography of the region and documented by a system of Pleistocene and Holocene faults that cut late Tertiary volcanic and sedimentary rocks, Quaternary sediments, and shorelines of the ancient pluvial Lake Alvord.

More significantly, the Alvord basin lies along the transition between the Great Basin extensional province and the Blue Mountains of Oregon and it is seismically quiescent. The youngest activity on the faults in the basin is on the order of 4 to 2 thousands of years old (kilo-annum, or ka). As such, the contemporary displacement field for the region is not affected by short-term displacement transients arising from earlier earthquakes. Based on continuous global positioning system (GPS) sites of the Pacific Northwest Geodetic Array, a 50-kilometer (km)-long baseline across the basin indicates west-northwest differential motion of 1.75 ± 0.3 millimeters per year (mm/yr).

The northerly trending Alvord basin is bound and internally deformed by an array of normal faults that, based on our extensive geophysical imaging, have relatively steep dips (~60º) to depths of 1.5 to 2.0 km. The faults have well-exposed scarps that cut Quaternary sediments and offset a system of shorelines of ancient Lake Alvord. Based on mapping using conventional digital elevation models and digital orthophoto-quadrangles as well as high-resolution surface imaging using a terrestrial laser scanner (TLS, or ground-based LiDAR) controlled by GPS positioning, we have identified two sets of shoreline terraces. The lower, younger set of 5 shorelines have ages of about 17 to 12 ka; the higher, older set of 3 shorelines have ages between to 350 to 130 ka. Along an east-west transect across part of the basin fault system, we used TLS to determine precise vertical tectonic displacements (cumulatively and at intervals) of shorelines across faults.

The highstand of the younger series and older series ofterraces is offset by 48.4±0.5 meters (m) and 80.1±0.6 m, respectively. The high-resolution digital elevation models produced from the TLS images confirm the remarkable geomorphic (elevational) stability of wave-cut surfaces over distances of several kilometers. As such, they provide an unprecedented vertical datum for active fault displacements.

In addition to large cumulative displacement on several mapped faults, we are able to measure differential offsets of individual shorelines across the structures. One fault records that activity affected 7 of the 8 shorelines, whereas other faults were active sporadically during the development of different lake shorelines. We have identified 17 "displacement events" bracketed by various shorelines, and each of the displacement events, based on the magnitude of vertical displacement, probably reflects several earthquakes. The slip measurements on various strands of the Alvord fault system indicate that the locus of displacement across the central Alvord basin was concentrated on one fault, but they also indicate that a significant percentage (~35 percent) of the total offset was distributed among several other fault strands across the basin.

A comparison of the geodetic rates (1.75 ± 0.3 mm/yr) and displacement rates estimated from offset of the older and younger series of shorelines indicates significant variability over 10^4- and 10^5-year time scales. For comparison with horizontal geodetic rates, vertical displacement rates determined from the shorelines are converted to horizontal rates by using the geophysically determined dip of the faults (60°). The younger series of shorelines (127 to 17 ka) were deformed prior to 4 to 2 ka, over an interval of 8,000 to 15,000 years, and provide a rate of up to 3.5 mm/yr, nearly twice the geodetic rate. In contrast, the estimated horizontal rate for the older series of shorelines (350 to 130 ka) range from 0.13 to 0.36 mm/yr, nearly an order of magnitude slower than geodetic rates.

The implications of these observations are (1) that faults in and around the Alvord basin experienced an elevated period of displacement rate (earthquake cluster) in the latest Pleistocene to middle Holocene; (2) average displacement rates since the latest Pleistocene are consistent with geodetic rates, which suggests that far-field displacement and strain accumulation are constant for the fault system over the last 10^4 years; (3) reduced rates for late Pleistocene shorelines suggest either a slower far-field displacement rate or that the locus of deformation moved within the basin and (or) to locales outside of the basin fault system; and (4) that the concentration of earthquake surface ruptures on individual strands of the Alvord fault system together with the distributed nature of cumulative displacement across the basin suggests that strain release and recurrence interval for earthquakes vary spatially within the fault system.

New Approach to Designing Geochemical Databases

By Branko Djapic[1], Sri Vinayagamoothy[1], and Kerstin Lehnert[1]

[1]Columbia University, New York City, N.Y.

Based on the success of the Petrological Database of the Ocean Floor (PetDB), other members of the geochemistry community have shown interest in building similar "infrastructure" databases for their disciplines. Specific requirements have added new complexity to the already complex data model, which required thorough examination of the basic premises on which the PetDB data model was built.

We have developed a new data model based on the requirements for SedDB (a database for sediment geochemistry) and VentDB (a database for the geochemistry of hydrothermal springs). The new data model is a natural extension of the PetDB model. It is designed in such a way that databases such as PetDB, SedDB, and VentDB become special cases of the generic model. This should allow for easier integration of various geochemical databases into a unified system.

We look at this data model as a multidimensional cube with the observed value as the basic building block. We have identified five basic independent components in the data model, each of which is multidimensional in itself: data source, sample, observed item, measuring point, and method. These dimensions have relations, which depend on the interpretation of the model. Irrespective of interpretation, the model should provide for an easy top-down application of metadata and corrections.

Another important issue for the database is the grouping of values that should be looked at together. This portion of the database is called "GeoModel," and the set of associated values is called "GeoModelContent."

New data can be derived based on GeoModel, which makes it a method for generation of new data. This way, both actual and derived data can be stored together in a fully integrated model.

We hope that an understanding of the basic factors involved, by scientists and database designers, will bring more commonality and interoperability than any database schema developed for specific purposes.

Session 3—Natural Hazards

Chair: Michael Williams, University of Massachusetts, Amherst, Mass.

Using the SCEC Computational Platforms for Seismic Hazard Analysis Research

By Philip Maechling[1], Thomas Jordan[2], J. Bernard Minster[3], Kim Olsen[4], Steven Day[4], Ralph Archuleta[5], Jacobo Bielak[6], David O'Hallaron[6], David Okaya[2], Ned Field[7], Hunter Francoeur[2], Joanna Muench[8], Scott Callaghan[2], Nitin Gupta[1], Vipin Gupta[1], Yifeng Cui[9], and the SCEC/CME Collaboration

[1]Southern California Earthquake Center, University of Southern California, Los Angeles, Calif.

[2]Department of Earth Sciences, University of Southern California, Los Angeles, Calif.

[3]Scripps Institute of Oceanography, University of California—San Diego, La Jolla, Calif.

[4]Department of Geological Sciences, San Diego State University, San Diego, Calif.

[5]Department of Earth Science, University of California—Santa Barbara, Santa Barbara, Calif.

[6]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, Pa.

[7]U.S. Geological Survey, Pasadena, Calif.

[8]Incorporated Research Institutions for Seismology (IRIS) Data Management Center, Seattle, Wash.

[9]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

Researchers working with the Southern California Earthquake Center (SCEC), a research consortium funded by the National Science Foundation (NSF) and the U.S. Geological Survey (USGS), are performing computationally intensive seismic-hazard research with the goal of better understanding earthquake processes and reducing earthquake hazards. To support SCEC research, geoscientists and computer scientists working on the SCEC Community Modeling Environment (SCEC/CME) Project have developed a highly capable computational environment that includes high-performance computing, large-scale data management, grid-based middleware, scientific visualization tools, and collaboration tools.

SCEC researchers have used the SCEC/CME computational environment to perform several large-scale simulations including wave propagation simulations (TeraShake), dynamic rupture simulations (TeraShake 2), and three-dimensional (3D) waveform-based, probabilistic seismic-hazard curve calculations (CyberShake). Our experience using the SCEC/CME system to perform seismic-hazard research has led us to develop a concept, which we call a "computational platform." A computational platform is defined as a vertically integrated collection of hardware, software, and people that provides a broadly useful research capability. The computational capabilities of a platform are validated and reusable. Computational platforms help to preserve vertically integrated computational research capabilities developed on large-scale cyberinfrastructure-intensive research projects.

We believe that the SCEC computational platform concept helps to capture several important and often overlooked aspects of scientific cyberinfrastructure. For example, the key to effective cyberinfrastructure for scientific research is integration between cyberinfrastructure specialties. Research groups like SCEC that use a wide variety of cyberinfrastructure must integrate diverse cyberinfrastructure specialties in order to perform research. The computational platforms preserve research capabilities after a significant amount of cyberinfrastructure integration has been successfully accomplished.

Another idea that is essential to the SCEC computational platform concept is the significance of highly skilled people in computational science research. Many software and tool development efforts, including knowledge capture and semantic reasoning efforts, are trying to make it easier for researchers to use the powerful and complex computational tools now available. However, in the computational platform concept, we acknowledge the essential role of knowledgeable people, including computer scientists, in computational-intensive science. The implications of this are that the geosciences must find ways to attract and retain computer scientists to support geoscience research efforts.

In this presentation, we will describe the SCEC computational platform concept and the existing SCEC platforms (including the OpenSHA, TeraShake, and CyberShake platforms) will be discussed. We will also outline some of the seismic-hazard research that has been performed using these tools.

Using GIS to Visualize, Analyze, and Forecast Seismic Hazards

By Serkan B. Bozkurt[1]

[1]U.S. Geological Survey, Menlo Park, Calif.

As part of a multi-institution earthquake-probability investigation of greater Tokyo, Japan, we are visualizing and analyzing the complex three-dimensional (3D) environment of Tokyo using advanced geographic information system (GIS) technologies. The goal is to develop a comprehensive description and understanding of the earthquake threat to Tokyo. The principal investigators of the project are the U.S. Geological Survey (USGS) and the Active Fault Research Center (AIST) of the Geological Survey of Japan (part of the National Institute of Advanced Industrial Science and Technology); participants in Japan include the National Research Institute for Earth Science and Disaster Prevention (NIED) and the Geographical Survey Institute (GSI) in Tsukuba, and the Japan Meteorological Agency (JMA) in Tokyo. As a result of this project, 13 scientific studies, authored by 24 scientists, have been submitted to scientific journals.

In this project, we have developed a GIS database that includes high-resolution topography and bathymetry, space imagery, 300,000 earthquake hypocenters, 10,000 seismic-intensity observations during the last 400 years, 3D tectonic plate models, fault locations, volcano locations, global positioning system (GPS) observations, triangulation stations, and urban features such as administrative boundaries and population information. As a result of this GIS effort, we have discovered a dislodged fragment of the Pacific plate beneath Tokyo (Stein and others, in press), and have also developed a statistical model to forecast distribution of seismic shaking based on 400 years of intensity observations (Bozkurt and others, in press). GIS analysis and visualizations played a crucial role in both of these studies.

The GIS effort of this project will be highlighted by showing several two-dimensional and three-dimensional visualizations and animations. Additional information and publications related to this project can be found online at the Tokyo Project Website at http://sicarius.wr.usgs.gov/tokyo/.

References Cited

Bozkurt, S.B., Stein, R.S., and Toda, S., in press, Forecasting probabilistic seismic shaking for greater Tokyo from 400 years of intensity observations: Earthquake Spectra.

Stein, R.S., Toda, Shinji, Parsons, Tom, and Grunewald, Elliot, in press, A new probabilistic seismic hazard assessment for greater Tokyo: Philosophical Transactions of the Royal Society, A—Mathematical, Physical and Engineering Sciences.

A New System for Publishing and Archiving Volcano Hazard Information Statements Using the Common Alerting Protocol (CAP)

By Dina Y. Venezky[1] and Seth Snediger[2]

[1]U.S. Geological Survey, Menlo Park, Calif.

[2]Alaska Volcano Observatory, Alaska Division of Geological and Geophysical Surveys, Fairbanks, Alaska

The U.S. Geological Survey (USGS) is responsible for monitoring U.S. volcanoes and issuing “timely warnings” of potential volcanic hazards to emergency-management authorities and the public. The monitoring efforts are divided among five volcano observatories, each with its own Web site and format for publishing information statements and updates. Recently, the USGS adopted the Common Alerting Protocol (CAP), an Extensible Markup Language (XML) format for distributing hazard information. Because each observatory monitors different volcanoes that are in various states of unrest, a flexible system was needed to meet individual needs while moving towards a more common format.

The new Web-based alert system uses PHP:Hypertext Protocol (PHP) and MySQL to create different types of alert messages, submit the XML to the Disaster Management Interoperability Services (DMIS), publish the alerts on the observatory Web pages, publish Really Simple Syndication (RSS) feeds of the alerts, and archive the information. A database schema for volcano hazard information was created for any observatory without an existing one. A common volcano event vocabulary is being developed using the following terms to highlight information for each volcano: background levels, deformation, gas emission, seismicity, hydrothermal, and volcanic activity.

Developing a system to meet the varied needs of the observatories was the biggest challenge. When the project began, each observatory had its own system of alert levels and (or) color codes for volcanic activity. For example, the Yellowstone Volcano Observatory sent regular monthly updates about Yellowstone, regardless of activity level and without a color code, whereas the Alaska Volcano Observatory sent daily or more frequent status reports about any volcano in the Aleutian chain with an elevated aviation color code and weekly updates covering all 30 monitored Alaskan volcanoes.

The new alert system provides a simple form for each observatory based on the volcanoes they monitor and codes they use. Fields simplify the input and selection of all required CAP attributes including report type, event, description, message status, type, scope, urgency, severity, and certainty. A preview screen shows both a formatted release and the XML. Once the alert has been sent, it is published in several formats, including RSS feeds available in real time on our Web sites, and stored in the database. The next step will be to further simplify the code, migrate to the new common alert level and aviation color codes for volcanic activity once the observatories have adopted them, and continue to examine methods for a standard description format. Additional information may be found online at http://volcanoes.usgs.gov.

Utilizing GEON Cyberinfrastructure for United Nations Disaster Relief Efforts

By Mark Gahegan[1], Suha Ulgen[2], and Dogan Seber[3]

[1]GeoVISTA (Geographic Visualization Science, Technology, and Applications) Center, The Pennsylvania State University, University Park, Pa.

[2]Field Information Support Project, Office for the Coordination of Humanitarian Affairs, United Nations, New York, N.Y.

[3]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

This talk introduces some possibilities for collaboration between the Geosciences Network (GEON) and the United Nations (U.N.) Office for the Coordination of Humanitarian Affairs, with a focus on supporting disaster-relief activities. One of the reasons why humanitarian disaster relief is often hampered is that it fails to use existing local expertise and data. The local community of scientists and experts who together hold vital knowledge and datasets are not usually equipped with any common technology or infrastructure, nor are they preformed into communities that are comfortable with collaborating.

GEON can contribute towards disaster preparedness in three important ways—

- Provide a uniform infrastructure and software stack to communities of scientists and other experts who have relevant expertise, which will facilitate collaboration.
- Encourage the development of communities created around GEON nodes that are deployed in developing countries.
- Host the UN datasets and software centrally and facilitate a U.N. GEON node, which can be deployed anywhere in the world at short notice.

A recent GEON workshop in India prior to deployment of a GEON Point of Presence (PoP) node highlighted the catalyzing effect such a node can have on the local science community—bringing geoscience agencies together, along with domain experts and computer scientists. If the development of such communities can be fostered in advance of a disaster in other countries, then the GEON cyberinfrastructure could form the basis of humanitarian relief efforts that can utilize local expertise and data, and that benefit from the established presence of a community of experts; it is too late to try and form one after a disaster strikes!

This talk will highlight recent work to develop an open-source geographic information system (GIS) software stack for UN disaster-relief mapping activities, discuss how existing GEON infrastructure could support such an effort, make some suggestions as to how local science communities might be better enabled, and open discussions as to how GEON and the U.N. might form a partnership to enhance the GEON Grid with further PoP nodes in developing countries.

Session 4—Web Services and IT Tools

Chair: Chaitan Baru, San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

A Virtual Geospatial Product Toolkit Based on Geospatial Web Service Orchestration

By Peisheng Zhao[1], Liping Di[1], and Yaxing Wei[1]

[1]George Mason University, Center for Spatial Information Science and Systems (CSISS), Laboratory for Advanced Information Technology and Standards (LAITS), Greenbelt, Md.

Typically, a geospatial processing model (or geoprocessing model) representing the domain knowledge involves a complex computing process. The execution of the model can generate a desired product; therefore, the model itself represents a geospatial product from which the virtual product can be generated on demand. It has been long recognized that the interoperability of spatial data and systems is critical to geospatial sciences. With the maturation of Web service technologies in recent years, a growing number of geospatial Web services have emerged to deal with spatial information in an interoperable way throughout the Web. Geospatial Web services are changing the way in which spatial information systems and applications are designed, developed, and deployed. With the Web service technology, a geoprocessing model can be implemented by assembling a chain of individual geospatial Web services. Because geospatial models embed lots of geospatial information and knowledge, domain-specific geospatial tools are needed to help users discover, retrieve, chain, and orchestrate these services. We have developed a virtual product toolkit that allows users composing Web service chains in a semi-automatic and dynamic way to represent virtual geospatial products. This paper illustrates the whole life cycle of virtual geospatial products based on Web service orchestration as follows: (1) discover relevant service type in Open Geospatial Consortium (OGC) Catalog Service for Web (CSW), (2) present and select matching service type at each step of model building with the assistance of geospatial ontology, (3) compose and validate the model based on domain knowledge, (4) instantiate the model into a concrete workflow in Business Process Execution Language (BPEL), and (5) execute the composed service chain automatically through its BPEL and Web Services Description Language (WSDL) scripts. All components of this toolkit are described in this paper. Catalog plays a "directory" role in this open, distributed Web service environment. Service providers register their services and related information using metadata, and service consumers can then query the metadata to discover information of interest to them. We have implemented the OGC CSW in this toolkit to help users either register or find useful services and related information. A graphic user interface in Model Designer helps users who are building geoprocessing models by dragging and dropping available components discovered in the CSW. BPELPower supports the dynamic execution of the BPEL-based Web service chain with various service invocation methods (for example, HTTP POST/GET, SOAP document/rpc). The Model Instantiation Service is used for translating the model into a concrete service chain in BPEL dynamically by collaborating with CSW. The Virtual Product Service plays a "control" role in the virtual product generation and provides virtual products to users through a standard interface, such as OGC Web Coverage Service.

Evaluating BDI Agents to Integrate Resources Over Cyberinfrastructure

By Leonardo Salayandía[1], Paulo Pinheiro da Silva[1], and Ann Gates[1]

[1]Department of Computer Science, University of Texas—El Paso, El Paso, Tex.

The vision of the Geosciences Network (GEON) is to enhance interdisciplinary collaboration of earth science efforts through the creation of a cyberinfrastructure (CI). The basic problem of enabling users to easily build complex CI-based applications and services by integrating resources available at GEON has been addressed by workflow technologies such as Kepler. This work aims to evaluate a Belief-Desire-Intention (BDI)-based agent as an alternative approach to Kepler workflows. The hypothesis is that BDI agents can adapt and better react to new, changing CI situations while executing workflows. BDI agents contain (1) a set of beliefs that reflect the current state of the world (for example, those provided by CI resources), (2) a set of desires that represent the state of the world that the scientist wants to achieve, and (3) a set of procedures that can be executed by the agent to achieve stated desires. The BDI agent instantiates procedures that are applicable to current beliefs (that is, intentions are created and executed by the agent). Additionally, the agent has the ability to backtrack to alternative intentions as beliefs are changed dynamically. In order to evaluate BDI agents to integrate resources over the CI, a simple geophysical procedure that utilizes several GEON resources is implemented using SRI International's Procedural Agent Realization Kit (SPARK), a mature implementation of a BDI agent framework. The objective of this research is to evaluate the adaptability of BDI workflows by evaluating the effectiveness of BDI agents to leverage CI while considering their extensibility to address issues that are still pending in CI developments. Some of these issues are transferring data between distributed resources, interoperability of independently created resources, and user evaluation of the validity of applications and services built from CI resources. Finally, a contrast is provided between workflow technologies and BDI agents.

The GEON LiDAR Workflow—An Internet-Based Tool for the Distribution and Processing of LiDAR Point Cloud Data

By Christopher J. Crosby[1], Jeffrey Conner[1], Efrat Jaeger-Frank[2], J. Ramón Arrowsmith[1], Ashraf Memon[2], Viswanath Nandigam[2], Gilead Wurman[1], and Chaitan Baru[2]

[1]Department of Geological Sciences, Arizona State University, Tempe, Ariz.

[2]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

Distribution, interpolation, and analysis of large LiDAR (light distance and ranging, also known as ALSM (Airborne Laser Swath Mapping)) datasets push the computational limits of typical data distribution and processing systems. The high point density of LiDAR datasets makes grid interpolation difficult for geoscience users who lack the computing and software resources necessary to handle these massive volumes of data. We are using a geoinformatics approach to the distribution, interpolation, and analysis of LiDAR data that capitalizes on a cyberinfrastructure that is being developed as part of the Geosciences Network (GEON) project. Our approach uses a comprehensive workflow-based solution, the GEON LiDAR Workflow (GLW) (http://www.geongrid.org/science/lidar.html), which begins with a user-defined selection of a subset of raw point data and ends with the download and visualization of interpolated surfaces and derived products. The Kepler workflow management system allows us to modularize and generalize the procedure. It provides the freedom to easily plug in new processes, to use existing subworkflows, and to easily extend or modify the analysis.

In the GLW, the billions of points within a LiDAR dataset point cloud are hosted in a spatially indexed database (IBM's DB2) running on the DataStar terascale supercomputer at the San Diego Supercomputer Center; this supercomputer is designed specifically for data-intensive computations. Data selection is performed via a Web map service (WMS)-based interface that allows users to execute spatial and attribute subset queries on the larger dataset. The subset of data is then passed to a Geographic Resources Analysis Support System (GRASS) open-source GIS-based Web service that handles interpolation to a grid and analysis of the data. The interpolation and analysis portion of the workflow was developed entirely within the open-source domain and offers spline with regularized smoothing and tension (rst) interpolation to grid with user-defined grid (digital elevation model, or DEM) resolution, as well as control over the spline parameters. We also compute geomorphic metrics, such as slope, curvature, and aspect as derived products from the DEM. Users may choose to download their results in either Environmental Systems Research Institute (ESRI) or American Standard Code for Information Interchange (ASCII) grid format as well as GeoTIFFs (TIFF files with geographic or cartographic data embedded as tags). Additionally, the GLW feeds into GEON Web services that are in development, which allows visualization of outputs in a Web browser window, or in three dimensions through (1) Interactive Visualizations Systems' (IVS's) iView3D viewer (for Fledermaus file formats), or (2) our own OpenGL-based tool, LViz (information is online at http://activetectonics.la.asu.edu/GEONatASU/LViz.html). Although the GLW was conceived for LiDAR data distribution and processing, most of the functions within this workflow are not limited to LiDAR data and may be used for distributing, interpolating, and visualizing any computationally intensive point dataset (such as gravity).

Currently, the GLW hosts three pilot datasets, which amount to a total of approximately 2.5 billion LiDAR returns. Ultimately, we believe that the GLW could be adopted as a valuable infrastructure resource for democratizing access to future LiDAR point cloud datasets for the geoscience community.

Efficient Search Index for Geospatial Data in a Relational Model

By G. Fekete[1], A. Szalay, and T. Budavari[1]

[1]Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, Md.

We discuss a project to develop a system for rapid data storage and retrieval using the Hierarchical Triangular Mesh (HTM) to perform fast indexing over a spherical spatial domain in order to accelerate storing and finding data over the Earth and sky.

Spatial searches over the sky are the most frequent queries on astrophysics data, and as such are central to the National Virtual Observatory (NVO) effort and beyond. The implementation of an HTM-based system has been highly successful; it produced a very fast search engine now used by at least 20 institutions worldwide, such as the Space Telescope Science Institute, National Aeronautics and Space Administration's Infrared Processing and Analysis Center, Royal Observatory Edinburgh, Institute of Astronomy at University of Cambridge, Sloan Digital Sky Survey, University of California—San Diego, Leiden Observatory, and University of Groningen. The library has applications in astronomy and earth science.

The goal is to speed up a query that involves an object (observation, location, and so forth) and a region of interest of an arbitrary shape (political boundary, satellite track, and so forth). In a very large database of hundreds of millions of objects, one wants to minimize the number of calculations needed to decide whether or not an object meets a spatial search criterion. We use a HTM index-based method to build a coarse representation of a cover map on the fly of the query region, which is then used to eliminate most of the objects that are clearly outside the region. False positives that pass the coarse test are then eliminated with more precise calculations.

One of the challenging problems—cross matching—is to find data on the same object in separate archives. Simple boxing of rectilinear constraints (that is, by right ascension and declination (RA,dec) or by latitude and longitude (lat/lon) limits) are inadequate because they are singular at the poles, unstable near them, and the actual shape of areas of interest do not always fit neatly within a box. For example, the track of an orbiting satellite with a great declination is a narrow strip such that the whole lat/lon-limited rectangle contains areas mostly outside the track. Furthermore, because of constraints imposed by instruments, engineering, and so on, scientists may need to define their own irregularly shaped query regions.

With the recent advances in the worldwide Virtual Observatory effort, we have now a standard Extensible markup Language (XML) data model for space-time data. This data model also provides a new standard way to express spherical polygons as search criteria.

Two planned future outcomes of this project are (1) a layer that enables our search engine to run inside relational standard language query (SQL) databases that are either Open Source (such as mySQL) or commercial (such as SQLServer), and (2) to participate as a first-class access method in relational database queries. The toolkit is implemented in a highly portable framework in C# programming language, which allows seamless integration with relational database engines and Web services, and in particular, makes it possible to develop a full Web-service implementation of the library that can be accessed through remote calls.

Session 5—Integration and Computation

Chair: June Thormodsgard, U.S. Geological Survey, Sioux Falls, S. Dak.

Exploring the Structure and Evolution of North America

By Gregory E. van der Vink[1] and Christel B. Hennet[1]

[1]EarthScope, Washington, D.C.

EarthScope is a national science initiative to explore the structure and evolution of the North American continent and to understand the physical processes controlling earthquakes and volcanoes. EarthScope is taking a comprehensive "systems" approach to investigate its scientific questions at all scales—from the active nucleation zone of earthquakes, to individual faults and volcanoes, to the deformation along the plate boundary, and to the structure of the continent and plate-tectonic motion.

EarthScope is unprecedented in its interdisciplinary approach to the geosciences and in its scope. With approximately $200 million in funding from the National Science Foundation's (NSF's) Major Research and Equipment and Facility Construction account, EarthScope will be developed over five years in partnership with the U.S. Geological Survey. Once completed, EarthScope is anticipated by the NSF to be operating for an additional 15 years, thus providing a primary source of data for the next generation of Earth scientists and educators.

To meet the project's scientific goals, EarthScope will drill a 3.2-kilometer-deep borehole into the San Andreas fault in California and install thousands of stations across the country. EarthScope's San Andreas Fault Observatory at Depth (SAFOD) is a comprehensive effort to drill and instrument a borehole through the San Andreas fault directly in the area of active earthquake generation. A suite of instruments in the borehole will directly measure the physical conditions under which these plate-boundary earthquakes occur. Covering North America and Alaska, EarthScope's network of 875 permanent global positioning system (GPS) stations and

100 borehole strainmeters will measure deformation across the active boundary between the Pacific and North American plates. EarthScope's USArray will install a dense array of 400 transportable seismic stations occupying 2,000 sites across the continental United States and Alaska to (1) record local, regional, and teleseismic earthquakes, and (2) produce high-resolution images of the Earth's interior. In addition, a pool of 100 GPS and 2,400 seismic campaign instruments will be available for temporary deployments.

Most of these stations will transmit data in real time via telemetry to data collection centers where the data will be freely and openly available to the scientific community, the educational community, and the public. EarthScope is expected to collect more than 70 terabytes (TB) of data during the first eight years of the project. Datasets will also include magnetotelluric data, airborne laser swath mapping, well logs, and fluid, gas, and core sample data. Scientists will integrate these data with a diversity of other geological data to address unresolved issues of the continental structure, evolution, and dynamics.

Building and Editing 3D Models of Lithospheric Structure Via Integrated Analysis of Geological and Geophysical Data—A Scheme for Formal Integration (Data Fusion) of Geophysical Data

By G. Randy Keller[1], Eva-Maria Rumpfhuber[1], Matthew Averill[1], Aaron Velasco[1], Vladik Kreinovich[1], and Ann Gates[1]

[1]Departments of Geological Sciences and Computer Science, University of Texas—El Paso, El Paso, Tex.

A fundamental goal of studying the lithosphere is to characterize the three-dimensional (3D) structure of a region of interest in order to address key scientific questions. This could be accomplished by determining physical properties, such as P-wave velocity (V_p), S-wave velocity (V_s), density, magnetic properties, electrical properties, anisotropy, attenuation (Q), temperature, and so forth for volume elements that could take several forms. In addition, interfaces that represent features such as the crust-mantle boundary (the Mohorovicic discontinuity, informally known as the Moho), the base of the lithosphere, the bottom of basins, faults, magmatic bodies, and so on must also be mapped in order to properly characterize a region. This goal can only be achieved through a highly integrated approach that takes advantage of all of the geological and geophysical constraints available; however, the building and editing of 3D models remains a serious challenge, as does the formal integration of differing types of data.

In most cases, seismic methods have the potential for providing the greatest resolution at depth but generally are the most costly approaches. Furthermore, many diverse techniques are available for each data type collected. Thus, developing an integration scheme for seismic results is an important first step in reaching the fundamental goal. Each type of seismic method has its own sensitivities and resolution, and when used alone, can constrain some aspects of the lithospheric structure; however, when used together with other types of data, the integrated dataset has the potential to significantly constrain fundamental aspects of the lithospheric structure and thus advance our understanding of the processes at work in the Earth, the natural hazards that result, and the natural resources that are produced.

Potential field data also can help constrain three-dimensional models of the lithosphere. For example, community-driven database and standardization efforts have ensured that gravity and magnetic data are available in all of North America. Because of the empirical relationship between density and seismic velocity, gravity data have long been employed at least as qualitative checks on seismic results, particularly in the lithosphere. Formal integration has been attempted in many ways over the years, but some recent 3D approaches are ideally suited for at least joint modeling, which may provide better overall models of lithospheric structure.

These approaches are a significant step towards comprehensively mapping 3D volumes. Also, iterating on the empirical relationship between seismic velocity and density provides a basis for joint analysis and for testing the effects of heat, fluids, and rock composition. Finally, the need for estimating and visualization of uncertainty in models is very important; this topic is addressed in a companion paper by Kreinovich and others (this volume).

CyberIntegrator—A Meta-Workflow System Designed for Solving Complex Scientific Problems Using Heterogeneous Tools

By Peter Bajcsy[1], Rob Kooper[1], Luigi Marini[1], Barbara Minsker[1], and Jim Myers[1]

[1]National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign, Urbana, Ill.

This paper addresses the problem of designing a highly interactive scientific meta-workflow system that aims at building complex problem-solving environments. The meta-workflow is viewed as a framework that integrates heterogeneous workflow engines, software tools, data sites, hardware resources, organizational boundaries, and (or) research domains. We present the architecture and implementation of a meta-workflow prototype called CyberIntegrator developed at the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. The need for CyberIntegrator stems from common geographic information system (GIS) problems amongst a number of ongoing observatory and disaster planning efforts, which are supported by cyberinfrastructures that are researched and developed at NCSA. Our current meta-workflow architecture enables users to (1) browse registries of data, tools, and computational resources; (2) create meta-workflows by example; (3) execute meta-workflows locally or remotely; (4) incorporate heterogeneous tools and link them transparently; and (5) provide recommendations about the usage of tools based on gathering and analyzing provenance information. The contributions of our work are in (1) defining the meta-workflow concept focused on science requirements and (2) building the technology and the prototype for CyberIntegrator software. We demonstrate the value of the CyberIntegrator software by applying it to two classes of complex problems: (1) data-driven analyses of multivariable relationships from remote-sensing data, and (2) Monte Carlo simulations of the maximum amount of pollution that a body of water can receive each day before it becomes unusable.

Towards Combining Probabilistic, Interval, Fuzzy Uncertainty, and Constraints—An Example Using the Inverse Problem in Geophysics

By V. Kreinovich[1], A.A. Velasco[1], M.G. Averill[1], R. Araiza[1], G.R. Keller[1], and G. Xiang[1]

[1]Departments of Computer Science and Geological Sciences, University of Texas—El Paso, El Paso, Tex.

Combining uncertainty models within cyberinfrastructure is a challenging problem. The main objective of cyberinfrastructure is to be able to seamlessly move data between different databases (where the data are stored in different formats), feed the combined data into a remotely located program (which may require yet another data format), and return the result to the user. It is also important to gauge the quality and accuracy of this result. We may have different models for describing the uncertainty of different databases and programs; it is, therefore, important to be able to combine different uncertainty models.

A Case Study—The Seismic Inverse Problem

The existing techniques for solving the seismic inverse problem, such as Hole's tomographic inversion code (Hole, 1992), only take into account the measurement results. Expert geophysicists also have some prior knowledge about the structure of the region. If we could take this prior knowledge into account when processing the geophysical measurements, then the accuracy of the geophysical structures uncovered by solving the inverse problem would be greatly improved, and also the convergence of the inversion would be aided.

Probabilistic Prior Knowledge

Measurement results often come with probabilistic uncertainty (that is, the probabilities of different values of measurement errors). Processing such measurement results yields (1) the estimate of the desired quantity, and (2) the standard deviation (and other statistical characteristics) of possible differences between the actual and estimated values of the desired quantity. Some prior knowledge comes from previous measurement processing and has, therefore, the form "the value *x* is approximately equal to *a*, with standard deviation *sigma*." This prior knowledge has been successfully used by Maceira and others (2005), among others.

The Need to Use Interval and Fuzzy Prior Knowledge

In some situations, there are different types of prior knowledge (for example, a geophysicist may know that the speed of sound at a certain depth must be between 6 and 8 kilometers per second (km/s)). We could, in principle, describe this information in probabilistic terms by assuming that 7 is the most probable value and 1 is the standard deviation, but this would be a distorted representation of the expert knowledge because, in contrast to this representation, the expert does not necessarily consider any value within the interval [6,8] to be more probable than others. Often, instead of a fixed interval, the expert can produce several nested intervals corresponding to different degrees of certainty—in effect, a fuzzy number. Techniques for handling such an interval and fuzzy uncertainty are described in Averill and others (2005).

The Need to Combine Different Types of Uncertainty

In real life, some prior knowledge comes from prior data processing and some from prior interval constraints. In this presentation, techniques that allow a combination of different types of uncertainty are described. The ultimate goal is to provide visual presentations of the combined effects of different types of uncertainty.

References Cited

Averill, M.G., Miller, K.C., Keller, G.R., Kreinovich, V., Araiza, R., and Starks, S.A., 2005, Using expert knowledge in solving the seismic inverse problem, *in* Proceedings of the 24th International Conference of the North American Fuzzy Information Processing Society, Ann Arbor, Mich., June 22-25, 2005: Piscataway, N.J., IEEE Press, p. 310-314.

Hole, J.A., 1992, Nonlinear high-resolution three-dimensional seismic travel time tomography: Journal of Geophysical Research, v. 97, no. B5, p. 6,553-6,562.

Maceira, M., Taylor, S.R., Ammon, C.J., Yang, X., and Velasco, A.A., 2005, High-resolution Rayleigh wave slowness tomography of Central Asia: Journal of Geophysical Research, v. 110, paper B60304, 15 p.

CHRONOS—Transforming Earth History Research by Seamlessly Integrating Stratigraphic Data and Tools

By Cinzia Cervato[1], Doug Fils[1] Geoff Bohling[2], Pat Diver[3], Doug Greer[4], Josh Reed[1], and Xiaoyun Tang[1]

[1]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

[2]Kansas Geological Survey, lawrence, Kansas

[3]DivDat Consulting, Wesley, Ark.

[4]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

Geologic time is the intellectual theme that connects a wide variety of research endeavors in geoscience—missing is the corresponding cyberinfrastructure that allows the resources of all these endeavors to be pooled. CHRONOS's purpose, emerging from community workshops dating back to 2001, is to transform Earth history research into a virtual online stratigraphic record by seamlessly integrating stratigraphic databases and tools. This online record accelerates and democratizes research by giving more users access to more data and by giving faster access to it. This presentation will give a broad overview of CHRONOS's mission and goals.

CHRONOS is a team of geoscientists and information technology specialists that is creating a cyberinfrastructure that delivers open access to a global federation of Earth history databases, tools, and services, thus providing the following:

- For academic, government, and industrial scientists—access to multiple, disparate databases on Earth history; data evaluation and conversion services; and powerful analytical tools.
- For autonomous databases, affiliated science initiatives, and data and tool contributors—a larger user community, greater visibility and acknowledgment, and access to tools and best practices, all without the cost and burden of reproducing interoperability.
- For educators, students, and policy makers—a convenient source of Earth history data, visualization tools, expert opinion, and educational materials.

The continued enhancement of the CHRONOS federation of resources will foster unprecedented cohesiveness within the geoscience community along with a growing pool of geoinformatics specialists. This will result in improved public awareness and appreciation of the knowledge that underpins the many hypotheses addressing contemporary issues such as climate change, evolution and extinction, and fossil fuel inventories.

Listening to countless discussions of the emerging system, CHRONOS members hear three pervasive practical concerns—convenience, quality control, and cost containment—from which some best practices emerge, as follows:

- CHRONOS's architecture is flexible, open, and searchable;
- Users have a choice of data-entry pathways;
- The linked databases maintain autonomy;
- CHRONOS integrates time-stratigraphic data;
- Tools and data services are developed together;
- Costs are distributed and shared with research opportunities.

CHRONOS's three organizing and operational initiatives provide the framework around which all of our tasks are managed. Each will be accomplished through the major activities identified with it, as follows:

- 1. Build cyberinfrastructure—Conceive, build, and support the robust system of integrated databases, tools, visualizations, and services that are essential to realize the CHRONOS vision;
- 2. Build and involve the CHRONOS community—Foster the growth of the CHRONOS community by actively promoting the open participation of diverse groups of users, data providers, and other geoinformatics efforts;
- 3. Give access to educational activities—Create an excitement for Earth history and geologic time through the implementation of human-computer interactions.

Semantically Enabled Scientific Data Integration

By Peter A. Fox[1], Deborah L. McGuinness[2], R. Raskin[3], and A. Krishna Sinha[4]

[1]High Altitude Observatory/Earth and Sun Systems Laboratory/National Center for Atmospheric Research, Boulder, Colo.

[2]Knowledge Systems, Artificial Intelligence Laboratory/Stanford University, and McGuinness Associates, Stanford, Calif.

[3]Jet Propulsion Laboratory, Pasadena, Calif.

[4]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Va.

The vast majority of explorations of the Earth system are limited in their ability to effectively explore the most important (and often most difficult) problems because they are forced to interconnect at the data-element (or syntactic) level rather than at a higher scientific (or semantic) level. In many cases, syntax-only interoperability is the state-of-the-art. In order for scientists and nonscientists to discover, access, and use data from unfamiliar sources, they are forced to learn details of the data schema, other people's naming schemes, and syntax decisions.

The Semantically-Enabled Scientific Data Integration (SESDI) project was created to demonstrate how ontologies implemented within existing distributed technology frameworks will provide the essential, re-useable, and robust support that is necessary for interdisciplinary scientific research activities. The project's goals are to enable an evolution toward science measurement processing systems (or frameworks) and to develop data and information systems (or frameworks) in support of specific applications. The initial focus of the project is on integrating volcanic activity data with atmospheric data in support of investigations into the relationships between volcanic activity and Earth's climate.

Another goal of this project is to provide scientists with the option of describing what they are searching for in terms that are meaningful and natural to them, instead of in a syntax that is not familiar. The missing elements in enabling higher-level interconnections have been the technology of ontologies, ontology-equipped tools, and semantically aware interfaces between science components that are made aware of the difference in semantics. In this presentation, the scope and some details of the current state of the project are covered. SESDI is funded by (1) the Advancing Collaborative Connections for Earth-Sun System Science (ACCESS) program, and (2) the Earth-Sun System Technology Office of the National Aeronautics and Space Administration (NASA).

CIG—Software for the Geophysics Community

By Walter Landry[1], Leif Strand[1], Luis Armendariz[1], Michael Gurnis[1], Michael Aivazis[1], Matthew Knepley[2], Charles Williams[3], and Luke Hodkinson[4]

[1]Computational Infrastructure for Geodynamics, California Institute of Technology, Pasadena, Calif.

[2]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill.

[3]Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, N.Y.

[4]Victorian Partnership for Advanced Computing, Carlton, South Victoria, Australia

Computational Infrastructure for Geodynamics (CIG) is a membership-governed organization which is managed by the California Institute of Technology (Caltech) for the National Science Foundation. CIG supports and promotes Earth science by providing state-of-the-art tools for computational geophysics using modern software development practices. CIG currently has 36 member institutions and 5 foreign affiliates. The goals of CIG are to (1) lead a coordinated effort to develop reusable, well-documented, and open-source geodynamics software; (2) construct the basic building blocks—an infrastructure layer—of software by which state-of-the-art modeling codes are quickly assembled; (3) create an extension of existing software frameworks to interlink multiple codes and data through a superstructure layer; (4) form strategic partnerships with the larger world of computational science and geoinformatics; and (5) develop specialized training and workshops for the geodynamics and larger Earth science communities.

CIG currently has working groups and active software development in several fields within geophysics: mantle convection, long- and short-term tectonics, and computational seismology. Working groups for magma dynamics and the geodynamo are still being formed. CIG is re-engineering existing codes and developing new software packages. In mantle convection, CIG is maintaining and distributing the latest version of the CitCom suite of finite element codes and is initiating the development of an entirely new code for compressible convection in a sphere. In long-term tectonics, CIG is developing GALE (Geodynamics Arbitrary Lagrangian Eulerian), a code for solving large three-dimensional problems of crustal evolution and basin formation. The short-term tectonics working group, in collaboration with a working group at the Southern California Earthquake Center (SCEC), is developing PyLith, a finite element code for pre- and post-seismic deformation and earthquake rupture dynamics. For the computational seismology community, CIG is re-engineering a suite of seismic-wave propagation codes, including SPECFEM, incorporating the Pyre superstructure framework. In addition, CIG is developing a suite of tools to compare modeling results, based on sets of community-developed benchmarks. These benchmarks are being integrated with the CIG software repository and build procedure. All CIG software can be found on the CIG Web site. Geophysics researchers are encouraged to participate in the CIG community, CIG-supported workshops and training sessions, and to visit the Web site at http://www.geodynamics.org to sign up for various mailing lists.

Session 6—Cyberinfrastructure and Visualization

Chair: Philip Maechling, Southern California Earthquake Center and University of Southern California, Los Angeles, Calif.

e-GEOS—eScience for the Earth- and Environmental Sciences

By J. Klump[1], P. Köhler[1], R. Häner[1], and J. Wächter[1]

[1]GeoForschungsZentrum Potsdam, Potsdam, Germany

Understanding Earth as a system is seen by the geosciences as fundamental tounderstanding the global challenges we face today, such as climate change or degradation of the natural environment. These challenges can only be met by interdisciplinary cooperation between the earth sciences and other branches of science and engineering. Modern information and communication technologies offer the chance to improve the scientific, technological, and management basis necessary to work on complex scientific questions and transfer the results into application. This new type of technology-assisted science is called eScience (enhanced science). eGEOS is the framework to build a generic eScience platform for the earth and environmental sciences, linking together geodata sources, geoscience libraries, and computing services. The platform will also supply a new generation of tools for the exploration and visualization of available information. These tools will help to create added value in research and development and support technology transfers into applications.

eGEOS offers an integrated approach to eScience. Building an eScience platform requires the development of technology and the availability of content, but it also requires to promoting a favorable science policy environment and establishing new ideas in knowledge and science management. This integrated approach assures a well-balanced process of transformation of the earth and environmental sciences into an eScience community. We can expect eGEOS to give us new tools for the utilization of information in the earth and environmental sciences. Examples are the (1) "Living Publication," which unites scientific publication with the publication of data; (2) "Geohazard Explorer," which manages and provides visualization of data on natural hazards; and (3) new services for exploring subsurface data, such as a standardized three-dimensional (3D) visualization tool. In addition to technical innovation, eGEOS will also facilitate the innovation of organizational models and process organizations which are aimed at advancing science management methods.

eGEOS is coordinated by GeoForschungsZentrum Potsdam (GFZ). The consortium presently has nine members. Alfred Wegener Institute for Polar and Marine Research (AWI) in Germany, the Centre for Environmental Research (Umweltforschungzentrum Leipzig-Halle), the German Geological Survey (BGR), and the Leibniz Institute of Marine Sciences (IFM-GEOMAR), together with GFZ Potsdam, combine many years of experience in the handling and processing of satellite, environmental, and lithosphere data. Essential contributions to eGEOS come from the participating libraries at the University of Göttingen, the Technical University School of Mines Freiberg, BGR, and GFZ. In addition, the consortium has considerable expertise in data and information management for the geological sciences. The consortium is completed by two industrial partners. The consultancy Spatial Business Integration GmbH has many years of experience in the optimization of research and business processes in geoinformation; lat/lon GmbH is an active player in the field of geospatial technology development and implementation. The project (which was formerly called GEO-GRID) is scheduled for a duration of three years.

The Geological and Ecological Sciences—Their Natural Interdependency Reflected in Cyberinfrastructure

By Susan D. Haseltine[1], Gladys A. Cotter[1], John P. Mosesso[1], and Annette L. Olson[1]

[1]U.S. Geological Survey, Reston, Va.

There are intricate linkages between the geosphere and biosphere, with life dependent on geological processes, and with life impacting the geosphere. The ecological science community is strongly dependent on geological data and information in studying ecosystems. An increased availability of geological and other types of data over the Internet, along with new analytical tools, will enable even deeper studies into complex ecological processes—and will help researchers make forecasts of environmental conditions. Access to geological data that are linked to specific locations, provide historical perspectives, and are available in real time is especially important in exploring rapid changes in ecological systems (such as those resulting from climate change and land use).

Despite strong challenges in serving diverse types of data from multiple disciplines, efforts are being made on many fronts to bring geological information to ecological researchers, and, in turn, to feed ecological data into geological models. Local, regional, national, and international collaborations are building networks and tools to share and serve information. Examples include the National Biological Information Infrastructure, a fully digital, interactive, and distributed system that helps enable the discovery, integration, and analysis of natural science data and information. The Gap Analysis Program (GAP) integrates biodiversity and land-cover data in order to provide conservation assessments of native species. New, innovative tools are being developed to harvest, integrate, and display information, such as the Columbia River USGS Interdisciplinary Science Explorer and the Global Data Toolset, each of which serve multidisciplinary data from distributed datasets on geospatial platforms. This sharing and integration of diverse data is supported by the application of standards and semantic tools. Even very diverse geological and biological data often have common ground through which connections can be made (for example, between geospatial locations and temporal frames).

The goal—and the challenge—is to build consistent information infrastructures (if not a common infrastructure) among scientific fields in order to support multidisciplinary research and decisionmaking. Furthermore, informatics must adapt to new data needs and to emerging technologies that are changing the way data are collected, stored, managed, disseminated, and applied. Potential benefits are tremendous: we can support complex research, provide more predictive power, make scientific data more valuable to more people, dramatically increase our understanding of the processes that shape our Earth, and provide greater scientific backing for complex resource-management decisions.

Scientific Information Management at the U.S. Geological Survey—Issues, Challenges, and a Collaborative Approach to Identifying and Applying Solutions

By David L. Govoni[1] and Thomas M. Gunther[1]

[1]U.S. Geological Survey, Reston, Va.

The U.S. Geological Survey (USGS) convened its first Scientific Information Management workshop in March of 2006. The workshop brought together scientists, information technology, and policy specialists to identify issues, to share ideas, and to showcase locally developed, but potentially broadly applicable, solutions. Science and technology specialists from all parts of the USGS agreed that most, if not all, key information management challenges cut across traditional discipline boundaries, and that they cannot be effectively addressed in a piecemeal fashion.

The need for a comprehensive information-management strategy requires that scientists and information professionals work together to ensure that the "system" of standards, processes, tools, and technology is appropriate, effective, consistent, well understood, and widely available throughout the organization. To assure that these goals are met, the USGS is making use of "communities of practice" as both a conceptual and practical approach to effective information sharing and problem solving. To date, more than a dozen such knowledge sharing communities have emerged, dealing with a range of topics. This presentation will outline a two-dimensional framework for understanding and addressing scientific information management issues in a scientific organization, discuss the importance of communities of practice in that framework, and show the USGS's first steps in these areas.

A Cognitive-Agent Based Geospatial Data Distribution System

By William G. Miller[1]

[1]U.S. Geological Survey, Reston, Va.

Geospatial information is usually contained in large datasets. There may be many simultaneous users of these datasets. Some of those users, like first responders to emergencies, may have priority needs. A distributed cognitive-agent architecture is proposed that allows individuals to self-associate into communities of interest and to share data with an access that is priority tuned for their community. Users can join and leave a community at will without disturbing the rest of the community. The architecture has a core infrastructure to provide for transactional updates and the synchronization of those updates with all subscribers. The core infrastructure also provides a deep archive and governance structure.

Designed using an Internet-overlay routing concept, the architecture provides for data neutrality, application-specific schema mapping, promiscuous caching (whereby data can be cached anywhere and at any time), data redundancy, and long-term persistence. Rights management is an integral part of the architecture because not all information will be available to the public, either because of its proprietary nature or security concerns.

The architecture supports client-server applications at the end nodes so that standards, such as those created by the Open Geospatial Consortium (OGC) and the International Standards Organization (ISO), are supported. The architecture assumes that requests are fulfilled to two main steps. First, a plan is created where the user can nominate multiple instances of resources to be utilized. A second phase executes the plan, optimizing the availability of resources at run time. The plans can be saved and reused.

Standard Compliant Grid Framework for Geoscience Virtual Product Modeling

By Aijun Chen[1], Yaxing Wei[1], Yuqi Bai[1], Liping Di[1], and Yang Liu[1]

[1]Center for Spatial Information Science and Systems, George Mason University, Greenbelt, Md.

Despite significant progress in the geosciences in acquiring, archiving, manipulating, analyzing, and using spatially explicit data for understanding physical, biological, and social systems on the Earth's surface, the on-the-fly production of customized and controllable, application-specific virtual products remains an issue. In this presentation, virtual products are defined as nonexistent data, information, knowledge, or decisionmaking advice that users need and can be produced through users' previous knowledge of existing data and services. The discussion is centered on two important issues: (1) the user must build up a virtual model according to the predefined geoscience abstract model in order to express their understanding for a specific application, and then they must convert the logical virtual model into a concrete, executable workflow based on available existing resources; and (2) users also must determine how to make the best use of such existing resources as data, services, computing, and storage in order to yield user-required data with high performance.

Grid computing has become the prevailing information infrastructure to facilitate the efficient and effective utilization of the computing, storage, data, services, and other resources dispersed on the Internet. The Open Geospatial Consortium (OGC) has developed a set of widely accepted Web-based interoperability standards and protocols for sharing geospatial resources for geoscience communities. Based on these two technologies, a geoscience-specific grid framework is proposed here to address the second point. First, we designed an implemented several geospatial grid services that follow the OGC geospatial Web services specification, such as Catalogue Service for Web (CS/W), Web Coverage Service (WCS), and Web Map Service (WMS). Second, the grid-enabled CS/W, WCS, and WMS are implemented on the framework not only for sharing computing capabilities, but also for archiving, managing, sharing, and serving geoscience data. Users may use these grid services to produce required geoscience products. An intelligent grid moderator service is proposed for intelligently moderating user requests and other grid service messages among geospatial grid services. Based on the grid CS/W service and grid replica location service, the grid optimization service is designed to improve the efficiency of data managing and serving. It is also used to select the best data and service instance when the concrete workflow is executed. The framework enables geoscience community users to model the virtual products and finally produce them through accessing and optimizing grid-managed distributed geospatial resources with OGC-standard interfaces without the need to either know or interface with underlying grid infrastructure.

Based on the framework, a prototype system demonstrates the visualized virtual modeling of a geoscience application, the optimized selection of data and services for instantiating the logical virtual model, and the production of performance-optimized virtual products. This research is supported by a the National Aeronautics and Space Administration (NASA)-sponsored project “Integration of OGC and Grid Technologies for Earth Science Modeling and Applications.”

Computer Vision and Digital Photogrammetry Methodologies for Extracting Information and Knowledge from Remotely Sensed Data

By Beata Csatho[1] and Toni Schenk[1]

[1]Civil and Environmental Engineering and Geodetic Science Department and Byrd Polar Research Center, The Ohio State University, Columbus, Ohio

An ever-increasing number of airborne and spaceborne data-acquisition missions with various sensors produces a glut of data. Sensory data rarely contain information in an explicit form that an application can directly use. The processing and analyzing of data becomes a real bottleneck; therefore, automating the processes of gaining useful information and knowledge from the raw data is of paramount interest. In this paper, we propose to employ the computer vision paradigm to extract information and knowledge as it pertains to a wide range of geoscience applications. The first part of the paper is concerned with introducing the paradigm, followed by the major steps to be undertaken for extracting information and knowledge from sensory input data. Features play an important role in this process. We describe the extraction of these features and their perceptual organization, provide examples, and follow this up with some examples of operations on features as they pertain to photogrammetry, remote sensing, and geographic information systems (GIS).

The second part of the paper demonstrates the feasibility of our proposed multidisciplinary approach for extracting information and knowledge from sensory input data. As an application example, we focus on the use of old photography for quantitative studies in the polar regions. An absolute prerequisite for the successful use of old photography is rigorous registration, either with other sensory input data or with respect to three-dimensional (3D) reference systems. Recent advances in digital photogrammetry allow registration with linear features, such as straight lines, curves, and free-form lines without the need for identifying identical points. The concept of sensor invariant features was developed to register such disparate datasets as aerial imagery and 3D laser point clouds that originate from satellite laser altimetry or airborne laser-scanning systems. After successful registration, digital orthophotos of old aerial photographs can be created for use in such studies as detecting changes in the landscape. Advanced pattern-recognition techniques (for example, object-based contextual image classification) greatly facilitate the delineation of different landscape units and the recognition and analysis of geomorphological features (for example, moraines, patterned ground, and volcanic structures) from multisensor datasets.

Next-Generation Geoscience Visualization Systems

By Brian Davis[1]

[1]EROS Data Center, U.S. Geological Survey, Sioux Falls, S. Dak.

A necessary companion to the computation of numerical models and large-scale image processing is the visualization of datasets and results. Researchers at the U.S. Geological Survey’s (USGS’s) Earth Resources Observation and Science (EROS) Data Center have collaborated with academic partners to address increasingly complex geoscience data-visualization requirements. The initial outcome of this collaboration was the GeoWall, which combined new projection technology with fast and inexpensive graphics cards and computers, thereby making it possible to provide a stereo projection system that is much more affordable than the previous commercial solutions. The GeoWall makes use of these components to view stereo imagery and three-dimensional (3D) models in a low-cost environment, thus enabling geoscientists to visualize the structure and dynamics of the Earth and help them understand spatial relationships. This computer-based implementation of familiar stereo concepts has the potential to bring virtual reality to a larger number of geoscience audiences. One unplanned outcome of the GeoWall collaboration was the formulation of the GeoWall Consortium, a group of researchers promoting the sharing of data, software, and information for the commodity-component virtual-reality community (http://GeoWall.org). Because of the momentum generated by the consortium, recent deployment of GeoWall systems has increased dramatically; the increase is partly because of the adoption of the technology by major geographic information system (GIS) software vendors, whose products are now compatible with GeoWall systems.

Although the GeoWall was a significant advancement in dealing with 3D data, it is not as well suited for high-resolution data. To address the need for high-resolution displays, the GeoWall-2 was conceived and developed in conjunction with academic partners. The GeoWall-2 is a scalable, cluster-based visualization system, comprised of varying numbers of computers and liquid crystal displays (LCDs). These systems enable a visualization of large quantities of remote-sensing data and imagery, map data, seismic interpretations, interactive museum exhibits, and other applications that require large collaborative screen areas. This technology has been developed by the National Science Foundation-sponsored OptIPuter project, which solicited the USGS to be a Federal affiliate partner. The goal of this next-generation cyberinfrustructure project is to develop new visualization architectures that will enable geoscience researchers, who are generating terabytes and petabytes of data, to interactively visualize, analyze, and correlate data from multiple storage sites connected to optical networks. A future generation of visual communication tools will be necessary so that (1) geoscience researchers, educators, and practitioners can interactively visualize the dynamic data that is generated by high-performance computing platforms, and (2) larger problem sizes, datasets, and geographic extents can be accommodated. This evolving technology is blurring the lines of distinction between computational and visualization systems, which must happen in order for remote collaborators to effectively communicate the results of their geoscientific research to policymakers, decisionmakers, and the public.

A GeoWall 3D display, a desktop stereo display, and a Personal GeoWall-2 (PG2) display will be described and demonstrated in order to portray the current state of the art in visualization systems.

Session 7—Oceans and Ice

Chair: Rama Kotra, U.S. Geological Survey, Reston, Va.

A Conversation With an Ocean Observing System

By Matthew Arrott[1], Luis Bermudez[2], and John Graybeal[2]

[1]University of California—San Diego, La Jolla, Calif.

[2]Monterey Bay Aquarium Research Institute, Moss Landing, Calif.

Many ongoing system-development efforts target an overarching architecture for interoperable ocean observing systems (IOOS). The IOO, and in particular the Ocean Observatories Initiative (OOI), are working on conceptual architectures for their respective cyberinfrastructures. Environmental observatories in other science fields are likewise developing architectures.

Meanwhile, dozens (if not hundreds) of marine science projects are developing actual software that can be adapted for, or adopted in, the yet-to-be-built systems. Projects and software developers are creating prototypes of service oriented solutions that can meet the anticipated needs of such interoperable systems.

Where and how will these design and implementation efforts come together? Informed by participation in architectural and implementation activities, some specific scenarios are being projected in which people, computers, and the cyberinfrastructure interact to achieve desired outcomes. There will be an outline of these scenarios at multiple levels: architectural concepts, concrete examples of scientific interactions, and (in a few cases) software examples. The applicability of these scenarios to different observing systems, such as OOI and IOOS, will be discussed.

Spatio-Temporal Asset Management in a Database of Glacier Change

By David Percy[1], Eric Hanson[2], Darrell Fuhriman[1], Cris Holm[2], William Garrick[2], Chris LeDoux[1], Matt Hoffman[1], and Andrew Fountain[1]

[1]Geology Department, Portland State University, Portland, Oreg.

[2]Academic and Research Computing, Portland State University, Portland, Oreg.

A comprehensive database of glacier extents for the western United States is being compiled in order to document and quantify glacier change during the 20th century. Sources for the database include printed historical maps, aerial and oblique photographs, digital data from the U.S. Geological Survey (USGS) 1:24,000-scale topographic maps, and satellite data from National Aeronautics and Space Administration (NASA). Scanned versions of all primary data sources will be included in the online interactive database.

In order to manage this vast set of imagery data and document the sources for all of our derived vector data, an “asset management” database was built that ties directly into our interactive Web-mapping interface. Each asset is uploaded to the Web server and tagged with variable metadata fields, depending on the asset type (oblique photograph, satellite image, and so forth). In order to manage the variability in attribute data, a hybrid resource description framework (RDF)-style database model was implemented. In this approach, a core set of data fields, including “asset_id,” are stored in a standard relational table structure. This is augmented with a table that contains only three fields: “asset_id,” property, and value. This “triples database” approach allows us to tag any data object with only the data relevant to its object type. This process simplifies queries because only one set of tables is needed; the alternative method is to build a separate relational table for every data type.

User interaction with the mapping interface allows for the selection of a hierarchy of assets associated with a specific area on the map managed through nested spatial hierarchies (such as state, glaciated region, or glacier). When the user is viewing the map, he or she may query it to show data about a specific glacier (such as its area, elevation, date mapped, and so on). In addition, the user is presented with links to the assets database, which in turn will show the specific assets associated with the specific spatial hierarchy selected.

The entire system is implemented using open-source tools for mapping, database management, and Web interfaces. These are Mapserver, PostGIS (an extension of PostgreSQL), PHP, Apache, and MySQL. Some of the front-end tools developed for this interface and data management application are intended to be released as open-source projects themselves.

Session 8—Hydrology

Chair: Pierre Glynn, U.S. Geological Survey, Reston, Va.

A Scalable System for Online Access of National and Local Repositories of Hydrologic Time Series

By Reza Wahadj[1], Ilya Zaslavsky[1], David Valentine[1], and David Maidment[2]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[2]Center for Research in Water Resources, University of Texas—Austin, Austin, Tex.

The Hydrologic Data Access System (HDAS) is a component of cyberinfrastructure for hydrologic research and education being developed within the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) hydrologic information system (HIS) project, which is funded by the National Science Foundation (NSF). Using Environmental Systems Research Institute's (ESRI's) ArcGIS Server and Microsoft's SQL Server, HDAS provides a visual map interface for browsing, searching, and visualizing hydrologic time series from multiple distributed sources of observation data.

In particular, HDAS facilitates user access to hydrologic time series from several federally maintained water-resource repositories, including (1) the U.S. Geological Survey's (USGS's) National Water Information System (NWIS), which provides streamflow data, ground-water measurements, and water- quality sampling; and (2) the Ameriflux tower network. Access to the various distributed data sources is provided via CUAHSI HIS Web services, which represent a common application programming interface (API) for querying metadata and time series stored in the hydrologic archives.

Several additional observation networks are being added, including (1) USGS's National Water-Quality Assessment (NAWQA) water-quality measurement network; (2) the Environmental Protection Agency's STORET (short for STOrage and RETrieval), which stores raw biological, chemical, and physical data on surface and ground water, (3) the National Oceanic and Atmospheric Administration National Weather Service's Automated Surface Observing System (ASOS), which is a network of meteorological variables; and (4) DAYMET's daily observations of temperature and precipitation. In addition, HDAS will provide access to hydrologic observation data collected by individual researchers.

HDAS is just one of the user interfaces that is being developed to access Web services within CUAHSI HIS. The open services-oriented architecture adopted by CUAHSI allows for HIS services to be accessed via popular desktop tools used by hydrologists (such as ArcGIS, Matlab, and Excel) and programming environments (such as Fortran, VB, and Java). This paper examines the trade-offs between uniform Web services design and accurate representation of individual hydrologic-data repositories, and describes how structural and semantic heterogeneities across hydrologic repositories have been resolved at each of the three levels of HIS: time-series representation and data modeling, Web services, and user-interface design.

HIS services follow the design of the Hydrologic Observation Database (HOD). HOD is distilled from multiple approaches to hydrologic observation databases by Horsburgh and others (2005). Metadata components of HOD-compliant databases are being generated now for the national repositories of hydrologic data. This information, including measurement stations and the types of measurement performed at each station, allows developers and users to program standardized requests against a particular repository. For example, a GetValues request takes a station ID, variable ID, and time range as inputs and returns a time series. HDAS uses such requests to display the measured values, visualize them as charts, download them as Microsoft Excel and CSV (comma-separated value) files, and optionally relay the data to an external time-series viewer.

HDAS is available at http://river.sdsc.edu/HDAS. More information about the CUAHSI HIS project is at http://www.cuahsi.org/his.

Reference Cited

Horsburgh, J.S., Tarboton, D.G., and Maidment, D.R., 2005, Hydrologic observations data, chapter 6 *in* Maidment, D.R., ed., Hydrologic Information System status report: Washington, D.C., Consortium of Universities for the Advancement of Hydrologic Science, p. 102-135, available online at http://www.cuahsi.org/his/docs/HISStatusSept15_ch6.pdf/. (Accessed September 7, 2006.)

Developing a CUAHSI HIS Data Node as Part of Cyberinfrastructure for the Hydrologic Sciences

By David Valentine[1], Reza Wahadj[1], Ilya Zaslavsky[1], David R. Maidment[2], and Tim Whiteaker[2]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[2]Center for Research in Water Resources, University of Texas—Austin, Austin, Tex.

The Consortium of Universities for the Advancement of Hydrologic Science, Inc., (CUAHSI) is working to implement a cyberinfrastructure for the hydrologic sciences. The CUAHSI hydrologic information system (HIS) is planned to be a live, multiscale, Web-portal system for accessing, querying, visualizing, and publishing water-observation data and models for any location or region in the United States. The architecture combines existing cyberinfrastructure from the

Geosciences Network (GEON) with commercial off-the-shelf geographic information system (GIS) software from Environmental Systems Research Institute (ESRI) and CUAHSI WaterOneFlow Web services to bring diverse, distributed data sources to the hydrologic sciences community.

At present, there are many Internet-accessible data sources that are used in hydrologic research (such as fluxes, flow paths, and mass balances). Many different methods are used to access, prepare, and process the available information. In order to create tools for the HIS, CUAHSI has implemented Web services which wrap existing Internet-accessible data. The Web services produce data in a standardized form for consumption in the hydrologic-observation data system. We call our application programming interface (API) "WaterOneFlow." Because WaterOneFlow is a standardized set of Web services, the API allows different services to be used by a variety of analysis clients including desktop applications (such as Microsoft Excel, MatLab, and ArcGIS) and a Web-based system such as the CUAHSI Hydrologic Data Access System (CUAHSI HDAS).

The infrastructure will consist of two parts: (1) a Microsoft Windows-based point of presence (POP) node, and (2) a GEON data node hosted by the San Diego Supercomputer Center (SDSC). The POP nodes will be distributed to various hydrologic observatories and will host a locally customized HDAS Web interface, WaterOneFlow services for access to the hydrologic observations data, and geospatial databases. The Windows-based node was a request from the community. The community uses a data model for ESRI's ArcGIS and ArcHydro, and has a familiarity with ESRI products. The Windows-based POP node software stack will include ESRI's ArcGIS Server software operating over HOD (Hydrologic Observation Datamodel) databases, which contain the observation metadata and data stored in MS SQL Server, and geoprocessing Web services. The POP nodes will utilize the GEON data nodes to catalog, distribute, and store information. Using the GEON cyberinfrastructure will allow for sharing between sites and disciplines. We can leverage the GEON to archive more than geospatial data, and to allow for the cataloging and retrieval of information that is not normally used in a commercial GIS environment. In addition, GEON provides digital library capabilities, long-term archiving of the information, and uses grid infrastructure processing and services. More details about CUAHSI can be found at http://www.cuahsi.org/. A report detailing the planning and status of CUAHSI HIS is available at http://www.cuahsi.org/docs/HISStatusSept15.pdf/.

Demonstrations

Three-Dimensional Model of Laccolith Intrusions in the Rosillos Mountains of Big Bend National Park—A Product of the Integration of Publicly Available Geospatial Data

By Eric D. Anderson[1]

[1]U.S. Geological Survey, Denver, Colo.

Publicly available geospatial datasets downloaded from the Web provide the basis for generating a three-dimensional (3D) geologic model of Tertiary igneous intrusions in the Rosillos Mountains of Big Bend National Park, Tex. The Rosillos Mountains have been mapped (using traditional field mapping techniques) as two stacked granitic laccolith bodies separated by a layer of limestone and sitting on Cretaceous strata. The main structural feature is the Tertiary northwest-to-southeast-dipping normal Chalk Draw fault, which cuts the eastern edge of the mountains and is clearly exposed to the north of the Rosillos Mountains. Traditional field mapping coverage is limited because of the rugged terrain; therefore, remote-sensing data and existing geospatial data are ideal for supporting the field geologic mapping of the area.

Datasets that help map the surficial geology include standard geologic mapping tools such as existing geologic maps, Landsat (Thematic Mapper) data, and digital elevation models (DEMs). Another surficial mapping tool is radiometric data, which, when plotted on a map base, help constrain surficial units by differentiating the chemical properties of the radioactive elements in each unit. These data provide information relevant to the understanding of various geochemical and physical processes, including alteration and weathering. The processed radiometric data from the Rosillos Mountains have helped to delineate both the lateral extent of and the chemical variations within the exposed intrusive bodies as well as the distributions of sediments derived from the mountains.

In addition to aiding the mapping of the lateral extents of the intrusions, the datasets that provide depth information (which allow extrapolation of the surficial geology to depth) also include aeromagnetic and gravity data. The aeromagnetic and gravimetric data have been processed and filtered to determine the lateral extent of the intrusions and locations of faults. Both two-dimensional (2D) and 3D forward modeling of the data have determined fault offsets and the subsurface extents of the intrusive bodies.

The interpretations of all processed datasets were georeferenced and integrated by using ArcGIS and EarthVision software. The digital integration has led to the development of a 3D model of the Rosillos Mountains. The 3D visualization of the geology in EarthVision facilitates interpretation of the tectonic and intrusive history of the Rosillos Mountains.

Volcano Monitoring Using Google Earth

By John E. Bailey[1,2] and Jonathan Dehn[2]

[1]Arctic Region Supercomputing Center, University of Alaska, Fairbanks, Alaska

[2]Alaska Volcano Observatory, University of Alaska, Fairbanks, Alaska

Modern advances in computing have led to a proliferation of three-dimensional (3D) image browsers. Although 3D visualizations of remote-sensing images have been possible for many years using techniques such as GIS, the level of complexity needed to tour these landscapes made it a cumbersome process even for those with technical expertise. One of the current leaders in 3D image-browsing capability, and the most successful in providing a worldwide distribution, is the Google/Keyhole product, Google Earth (GE). With its established clientele base, worldwide name recognition, and superior user-friendly interface, GE is leading the way for geobrowsers.

Google Earth utilizes KML (Keyhole Markup Language), which is becoming the geobrowser standard; a KML importer will be included in the National Aeronautics and Space Administration's (NASA's) WorldWind and Environmental Systems Research Institute (ESRI) plans to include similar capability in its upcoming ArcGIS Explorer. Although criticized by some sections of the media as fun to look at but of no practical value, geobrowsers provide more than just an ability to view or fly through visual images draped over an elevation model. They offer an interface for integration and visualization of both images and other types of geographical data, which makes them potentially powerful tools in the realm of earth sciences.

At the Alaska Volcano Observatory (AVO), Google Earth is being used as an interface for visualizations of the 2006 eruption of Augustine Volcano. The continuing development of GE is producing interactive, real-time-monitoring capabilities for remote sensing of volcanoes. Examples include the ability to (1) browse Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) image overlays with dynamic control, (2) view Webcams through the GE browser, (3) retrieve location and status information on instrumentation, and (4) animate the modeled 3D ash plumes. Currently, AVO is developing a method to integrate satellite-based thermal-anomaly- and ash-plume-detection alarms.

NASA's Earth Science Gateway—An Open Standards-Based Portal to Geospatial Data and Services

By Myra J. Bambacus[1], Nadine Alameh[2], and John Evans[3]

[1]GeoSciences Interoperability Office, National Aeronautics and Space Administration (NASA), Goddard Space Flight Center, Greenbelt, Md.

[2]MobiLaps LLC, Silver Spring, Md.

[3]Global Science and Technology, Inc., Greenbelt, Md.

NASA's Earth Science Gateway (ESG) streamlines access to remote geospatial data, imagery, models, and visualizations through open, standard Web protocols. By organizing detailed metadata about online resources into a flexible, searchable registry, it lets scientists, decisionmakers, and others access a wide variety of observations and predictions from NASA and other sources of natural and human phenomena related to earth science and the Earth-Sun System.

This computer demonstration is designed to take the users on a tour, starting with browsing the gateway for data, discovering relevant data, and then directly connecting to the underlying sources via interoperable interfaces. Finally, the gateway is used to visualize and overlay the data in an intuitive user interface. The demonstration will also view the ESG from a data provider's perspective by following the publication of the metadata and online resources through the gateway. We hope to demonstrate that data providers can make their resources more easily accessible to traditional and nontraditional users.

What Do We Know? Tomography Uncertainty Visualization

By Gregory Bensen[1]

[1]University NAVSTAR (Navigation Signal Timing and Ranging) Consortium (UNAVCO), University of Colorado, Boulder, Colo.

We present a method for handling seismic-tomography data, which includes uncertainty and anisotropy visualization and a potential standard data description and storage format. In many publications of seismic-tomography models, at least cursory treatments of model uncertainties are presented. Different approaches have been tried to show this uncertainty both in two-dimensional (2D) plots and in three-dimensional (3D) representations. Similarly, anisotropy measurements can also be difficult to visualize. The model of Shapiro and Ritzwoller (2002) reports shear-wave velocities (v_s, v_{sv}, and v_{sh}) and corresponding minimum and maximum values resulting from a Monte-Carlo inversion. Examples of visualization of uncertainty and on-the-fly computing of anisotropy will be presented using the Integrated Data Viewer (IDV) in both 2D and 3D representations.

Using tomographic models from different researchers can be a difficult endeavor. The models are in widely varying formats and, even after tedious data reformatting, it is still often difficult to view models in anything but simple (north to south, or east to west) 2D slices. At UNAVCO, as part of the Geosciences Network (GEON) project, we have been working on establishing a group of hosted seismic tomography models. These models follow our standard minimum amount of descriptive information, such as authors' names, tomography technique, source publication, and so on. In addition, we have been making these data available in the network Common Data Form (netCDF) format, which is required for viewing in the IDV. NetCDF consists of (1) a set of interfaces for array-oriented data access, and (2) a freely distributed collection of data-access libraries. NetCDF is being used by many groups as a standard way to represent a range of scientific data and models. This format allows nonspecialist researchers to download and use seismic-tomography models in 3D, while clearly giving credit to the authors, stating the techniques that were used to construct the model, and conveying the constraints or limitations of the model.

Reference Cited

Shapiro, N.M., and Ritzwoller, M.H., 2002, Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle: Geophysical Journal International, v. 151, no. 1, p. 88-105

A Novel Search Mechanism for Navigating National Datasets

By Bora Beran[1] and Michael Piasecki[1]

[1]Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, Pa.

National datasets such as the U.S. Geological Survey's National Water Information System (NWIS) and the Environmental Protection Agency's (EPA's) Storage and Retrieval (STORET) system offer a considerable volume of valuable information for researchers and engineers. Even though these national datasets maintain consistency within themselves to some extent, it is a challenge to make them work together for several reasons; semantic heterogeneity and the level of granularity are the two most dominating issues. Also, in most data-access portals, the complexity of systems directly affects the user's experience; complexity can translate into dozens of Web forms and subsequent form submissions. Moreover, the availability of metadata is sporadic, so much so that retrieved datasets sometimes offer nothing more than column headers.

If we use the problem of getting the STORET and NWIS systems to work together as an example, we find that both systems assign different parameter codes and (or) names to the same observation, whereas some parameters do not have equivalents because of a difference in granularity. One example is that NWIS offers both "Ammonia Nitrogen (Bed Sediment)" and "Ammonia Nitrogen (Suspended Sediment)" measurements, but STORET offers only "Ammonia Nitrogen (Sediment)" measurements. Recent surveys conducted by two environmental observatory initiatives funded by the National Science Foundation—the Collaborative Large-Scale Engineering Analysis Network for Environmental Research (CLEANER) and the Consortium of Universities for the Advancement of Hydrologic Sciences, Inc., (CUAHSI)—clearly indicate that one of the most sought-after features a data access system should have is "…to make the data world more coherent and to allow access to all holdings through just one access point…."

In this demonstration, we'll present a system that will allow users to query several national datasets through a single, intuitive interface that will not only help them to easily pinpoint the datasets they were looking for, but will also query for several relevant parameters at once by entering a keyword, such as "pesticides."

MMI Web Service

By Luis Bermudez[1] and John Graybeal[1]

[1]Marine Metadata Interoperability, Monterey Bay Aquarium Research Institute, Moss Landing, Calif.

The Marine Metadata Interoperability (MMI) Project promotes the exchange, integration, and use of marine data through enhanced data publishing, discovery, documentation, and access. One of our goals is to provide interoperable demonstrations bringing together heterogeneous datasets. We will explain the design architecture of Tethys, which is one of these interoperable experiments. We will explain how semantic mediation services can be used by both data producers and data clients, and describe how these services may be implemented by both the producer and client.

Solid-Earth Science Computational Architecture and Environment for Space-Based Data Analysis and Modeling

By Ronald Blom[1], Margaret Glasscoe[1], Richard Gross[1], Eric Gurrola[1], Matthew Henderson[1], Kenneth Hurst[1], Erik Ivins[1], Jeffery Jewell[1], Daniel Katz[1], Jay Parker[1], Carol Raymond[1], Paul Rosen[1], Xiaoping Wu[1], and Cinzia Zuffada[1]

[1]Jet Propulsion Laboratory, Pasadena, Calif.

Detailed understanding of processes underlying solid-Earth deformation and mass transport requires a nontraditional, integrated, interdisciplinary approach that is dependent on multiple space- and ground-based datasets, modeling capabilities, and computational tools. At present, details of geophysical data acquisition, analysis, and modeling largely limit research to domain experts. Future interdisciplinary research can be enabled by computational architectures that are optimized to perform data processing of multiple data types in a user-friendly environment. We report progress on a Web-based computational framework with applications for specific data and geophysical modeling tasks that are focused on space-based datasets, including those derived from Interferometric Synthetic Aperture Radar (InSAR), Gravity Recovery And Climate Experiment (GRACE), and the Global Positioning System (GPS). This framework will include an integrated modeling capability for high-resolution deformation and gravity, forward models of visco-elastic mass loading over short wavelengths and complex time histories, forward-inverse codes for characterizing surface-loading responses over a range of time scales, and the inversion of combined magnetic and gravity fields to constrain deep-crust and mantle properties. The system uses a three-tier architecture; a middle tier server manages user projects, resources, and security and also enables future scalability. Users log into a Web page and control a personal project area that is maintained for them between sessions. Upon selection of an application and a host from a list of available entities, inputs may be uploaded or they can be constructed from Web forms and available data archives. Results are posted via Web page addresses (URLs). Interdisciplinary work is supported through the availability of all applications via browsers, application tutorials and reference guides, and completed examples. Demonstration problems related to current research guide development of the computational architecture. These problems include analysis of Gulf Coast subsidence, global ice-mass changes, gravity and deformation of a layered spherical-earth model due to large earthquakes, and an analysis of the rift setting of Lake Vostok in Antarctica.

National Archive of Marine Seismic Surveys—Status Report on a U.S. Geological Survey Program Providing Access to Proprietary Data

By Jonathan R. Childs[1] and Patrick E. Hart[1]

[1]U.S. Geological Survey, Menlo Park, Calif.

Seismic-reflection-profile data provide images of the Earth's upper crust to depths of tens of kilometers. These images are invaluable for stratigraphic and structural analyses that are used in assessing geologic hazards, resource potential, geologic framework, and other earth science issues.

During the last four decades, hundreds of thousands of line kilometers of two-dimensional marine seismic-reflection data have been collected by the hydrocarbon exploration industry within the United States' Exclusive Economic Zone (EEZ). The commercial value of much of these data has decreased significantly because of drilling moratoria and new technology such as three-dimensional data acquisition; however, these data still have tremendous value for scientific research and educational purposes. The U.S. Geological Survey (USGS) has recently entered into agreements with commercial owners of large data holdings to transfer to the public domain over 250,000 line kilometers of data from offshore the East, West, and Alaska coasts of the United States. In order to provide access to the data, the USGS has developed the National Archive of Marine Seismic Surveys (NAMSS) program. For a small fraction of the money that would be required to collect new data, work is underway to organize and reformat digital data currently stored on tens of thousands of nine-track tapes. The NAMSS Web site at http://walrus.wr.usgs.gov/NAMSS/ provides trackline maps of surveys that are now or will soon be available for download in a SEG Y format (an open standard developed by the Society of Exploration Geophysicists). As more data providers and users become aware of this new resource, it is hoped that additional partners will join this data rescue effort.

RIMS—A Real-Time, Interactive Mapping System for Neotectonic/Geologic Investigations of Digital Terrain Data

By Eric Cowgill[1], Adam Forte[1], Tony Bernardin[2], Oliver Kreylos[2], and Bernd Hamann[2]

[1]Department of Geology, University of California—Davis, Davis, Calif.

[2]Institute for Data Analysis and Visualization (IDAV), Department of Computer Science, University of California—Davis, Davis, Calif.

The current flood of terrain and multispectral satellite data with regional to planetary coverage is both overwhelming our analytical capabilities and producing fertile opportunities for discovery. As a result, we need new analytical tools that can handle this flood.

Although automated approaches (such as feature extraction) are very useful, they presume users can specify formal rules by which their data may be filtered. An alternative approach is to visualize the data and allow geologists to directly observe, measure, and record geometric patterns present within them via geomorphic and (or) geologic mapping. An important advantage of this latter process is that it allows users to adapt or evolve their selection criteria as the analysis proceeds and thereby discover previously unrecognized features or patterns. Unfortunately, this approach is generally limited by hardware capabilities because it requires that the data be displayed. Earth and planetary geologists have struggled with this limitation for some time and in particular with the difficult task of merging their detailed observations from large areas (1,000 X 1,000 km) directly with high-resolution (1- to 10-meter resolution) remotely sensed data.

In response to this problem, we have developed a real-time interactive mapping system (RIMS). RIMS uses a quadtree-based multiresolution method to circumvent hardware limitations and permit real-time rendering of high-resolution data over large spatial areas. The resulting visualization environment is analogous to Google Earth in the way that it allows users to merge terrain models consisting of texture data with a digital elevation model (DEM); this process essentially drapes the terrain model over the DEM without having to manipulate either dataset. More importantly, RIMS also allows users to map inside this interactive environment by generating vector-based polylines directly on the terrain visualization and by providing a "virtual geologic compass" for measuring orientations of geological surfaces. RIMS also provides sculpting tools for three-dimensional reconstruction of deformed and partially eroded surfaces, such as folded bedding. The georeferenced maps made using RIMS can be exported to standard geographic information system (GIS) software.

Using RIMS, we are currently mapping several zones of active intracontinental deformation, including the northern edge of the Arabia-Eurasia collision zone that is located between the Black and Caspian Seas. This area lies along the politically volatile northern foothills of the Greater Caucasus Mountains in Chechnya. Using Shuttle Radar Topography Mission 90-meter (SRTM 90) data and Landsat Enhanced Thematic Mapper Plus (ETM+) data, we are characterizing potential seismic sources to determine if the total shortening across the range increases systematically from west to east, as observed by Global Positioning System (GPS) measurements of the instantaneous deformation field.

Our RIMS-based mapping has revealed evidence of recent folding along two east-west-trending ridges within the foreland basin, subparallel to the Greater Caucasus Mountins to the south. A fluvial terrace along the northern edge of the northernmost ridge is both elevated relative to the active flood plain within the foreland basin and locally bowed into east-west-trending linear ridges. The terrace incision and secondary ridge formation most likely resulted from recent folding. At both the western and eastern tips of the northern ridge, this fluvial terrace grades upslope into a pediment surface that is locally breached by gullies draining the ridge, but is preserved as a set of triangular facets along the interfluves between the gullies. The local preservation of this relict surface attests to recent lateral propagation of the fold tips. Our Caucasus research demonstrates that the strength of RIMS is its ability to combine interactive rendering with the interactive mapping and measurement of features observed in topographic and texture data.

AmericaView—A State-Based Remote-Sensing Initiative in Partnership with the U.S. Geological Survey

By Rebecca L. Dodge[1] and Karen Zanter[2]

[1]GeorgiaView Consortium, Department of Geosciences, University of West Georgia, Carrollton, Ga.

[2]AmericaView Project, Earth Resources Observation and Science (EROS) Data Center, U.S. Geological Survey, Sioux Falls, S. Dak.

AmericaView (AV) is a nationwide program that was created to advance the availability, timely distribution, and widespread use of land-based remote-sensing data, especially to users in the university community. AmericaView is administered through a partnership between the U.S. Geological Survey (USGS) and the AmericaViewSM Consortium. The Consortium is the Federal Government's primary partner in achieving the AV program vision and goals, which focus on helping the university, secondary-education, and public sectors in each State identify, develop, and distribute the kinds of remote-sensing applications each State needs most. The consortium consists of university-led, State-based groups working together to build a nationwide network of State and local users through education, training, public outreach, and applied research.

On the USGS side of the partnership, the AV program is a data infrastructure, data processing, and data delivery initiative administered under the Land Remote Sensing (LRS) Program and carried out by the USGS Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, S. Dak. One of the obstacles that researchers encountered in the past was the difficulty of identifying and acquiring the most appropriate data for their particular application and study area. The AV program funded the development of the USGS Global Visualization (GloVis) viewer to overcome this barrier. GloVis added the capability to efficiently and graphically search the data inventories held at EROS. The holdings include land-based remote-sensing data from the Landsat, Earth Observing-1 (EO-1), Terra, and Aqua satellites. In addition, the software code for GloVis has been adapted by many of the AV participants to create similar visual Web-based tools to access individual State remote sensing data archives. Another tool prototype developed at EROS is an image acquisition viewer. This viewer allows users, such as educators or emergency response personnel, to use an Internet browser to view data acquisitions in near real time at the EROS ground station.

The need for AV has been building for more than 30 years. Since the early 1970s, the Federal Government and private sector have spent billions of dollars on satellite-based Earth-observation systems and have worked with the research community to identify, develop, and distribute real-world applications for mapping, monitoring, and managing natural and environmental resources. Unfortunately, development and distribution of real-world applications have persistently been tough issues for both the Federal Government and the academic research community. Satellite data are expensive, and using the data requires significant investments in software, hardware, and training. Because of the costs involved, it has often been hard for university researchers to use or even access the data, particularly at smaller schools or research facilities. For three decades, cost issues have hindered applied research and made it difficult to train the workforce, both current and future; cost issues also meant that many State and local agencies that work with applied research programs have not been able to effectively integrate remote-sensing technology into their management or decision-support programs. The consortium is charged with helping each State overcome these difficulties by establishing university-led State consortia and with facilitating interstate networking and mentoring to support technology transfer. The AmericaViewSM Consortium is actively working with the USGS and universities across the country to expand participation in the AV program to all 50 States.

The Virtual Solar-Terrestrial Observatory—Exploring Paradigms for Interdisciplinary Data-Driven Science

By P.A. Fox[1], D. Middleton[1], D.L. McGuiness[2], L. Cinquini[1], J.A. Darnell[1], J.L. Benedict[2], J. Garcia[1], P. West[1], and S.C. Solomon[1]

[1]High Altitude Observatory/Earth and Sun Systems Laboratory/National Center for Atmospheric Research, Boulder, Colo.

[2]Knowledge Systems, Artificial Intelligence Laboratory/Stanford University, and McGuinness Associates, Stanford, Calif.

Virtual observatories can provide vast stores of data concerning scientific observations. As these electronic stores become widely used, there is potential to improve the efficiency, interoperability, collaborative potential, and impact of a wide range of interdisciplinary scientific research activities. To realize this potential, technical challenges need to be addressed concerning (at least) representations and interoperability, access, and usability of data. The Virtual Solar-Terrestrial Observatory (VSTO) project is funded by the National Science Foundation (NSF) and is a joint effort between the High Altitude Observatory and the Scientific Computing Division at the National Center for Atmospheric Research (NCAR) and McGuinness Associates (a consulting firm). The VSTO project is beginning to provide an electronic repository of observational data that spans the solar-terrestrial physics domain. We are also designing and implementing tools and an infrastructure for accessing and using the data. Our main contributions include (1) building the repository, infrastructure, and tools for particular solar-terrestrial physics and (2) developing a design and infrastructure that may be broadened to cover more diverse science areas and user communities.

In this presentation, we describe the goals and present the design, the current and planned prototypes, and the technical infrastructure of the VSTO. We also discuss what we have learned about the processes involved in developing the VSTO, the required semantics, and how the semantics affect the framework architecture, choice of technologies, and service interfaces.

SCEC Earthworks—Community Access to Wave Propagation Simulations

By Hunter Francoeur[1], Joanna Muench[2], David Okaya[1], Yifeng Cui[3], Philip Maechling[1], Ewa Deelman[4], Gaurang Mehta[4], and Thomas Jordan[1]

[1]University of Southern California, Los Angeles, Calif.

[2]Incorporated Research Institutions for Seismology (IRIS) Data Management Center, Seattle, Wash.

[3]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[4]University of Southern California Information Sciences Institute, Marina Del Rey, Calif.

Researchers with the Southern California Earthquake Center (SCEC) Community Modeling Environment (CME) are developing the SCEC Earthworks Science Gateway that allows users to configure and execute earthquake-wave-propagation simulations through an easy-to-use, portal-based interface. The development of the gateway is motivated by the recognition that SCEC researchers want to use high-performance wave-propagation simulations in their research, but the knowledge needed to run high-performance codes and to use grid-based workflow tools presents a serious barrier to entry. The SCEC Earthworks system provides users with access to the power of a grid-based scientific workflow system and the computational and data management power of the TeraGrid, but it can be used without detailed knowledge of the computational environment and geophysical codes.

The SCEC Earthworks portal is based on the software and cyberinfrastructure developed by the SCEC/CME Project. SCEC Earthworks' wave-propagation simulations use well-validated geophysical models and high-performance simulation software. When wave-propagation simulations are run, the SCEC Earthworks system generates a well-defined set of data products, including synthetic seismograms and ground-motion maps. SCEC Earthworks also interfaces with the Incorporated Research Institutions in Seismology (IRIS) Data Handling Interface (DHI), which provides access to observed data that include earthquake catalogs and seismograms.

Users access the SCEC Earthworks system through a Web-based portal built using the GridSphere Portlets engine. Using a portlet-based interface, users can configure, submit, and monitor wave-propagation simulations. They can also access the resulting simulation data products. The portlets allow users to browse simulation data products, save configurations, and share simulation results with other users.

All steps in the wave propagation simulations (including mesh generation, wave propagation, and post processing) are run using a grid-based workflow system based on the Virtual Data System (VDS), the Pegasus metascheduler system, and the Globus toolkit. These workflow tools perform the backend steps of registering data with a replica location service (RLS) and building, submitting, and monitoring workflows. The metadata for the resulting data products are registered within a metadata catalog service (MCS). A Java library mediates the grid computing resources. Only authenticated users may run simulations, and the total simulation size is limited to conserve resources.

The SCEC Earthworks Science Gateway represents an important step towards wider SCEC community access to high-performance computing resources. The gateway allows members of the SCEC community to share metadata-driven seismic simulations and their resulting data products, thus harnessing the power of the TeraGrid to enable important research initiatives.

Managing and Utilizing Knowledge for GEON

By Mark Gahegan[1], Steve Weaver[1], and Tawan Banchuen[1]

[1]GeoVISTA (Geographic Visualization Science, Technology, and Applications) Center, Department of Geography, The Pennsylvania State University, University Park, Pa.

Separating data from knowledge always causes problems and is a common complaint about computer-based methods. The problem is exacerbated by cyberinfrastructure, which makes data all the more accessible. This talk describes our efforts to collect, represent, and use geoscientific knowledge more effectively by using the ConceptVista tool (http://www.geovista.psu.edu/ConceptVISTA/launch.jsp), which is part of the knowledge infrastructure under development by Geosciences Network (GEON) researchers.

ConceptVista is a concept mapping and ontology editing environment with some important distinctions: (1) it can reason with incomplete or ambiguous knowledge; (2) it can perform semantic searches via the Google, Amazon, and Digital Library for Earth System Education (DLESE) search interfaces and therefore find useful Web content, books, and educational resources; (3) it can link to CiteSeer (the Scientific Literature Digital Library created at the Pennsylvania State University) so it can also map out knowledge domains and social networks surrounding researchers; (4) it can link to remote resources, such as datasets, Web pages, and imagery, so it can be used as a front end to data collections; (5) it can support some complex strategies for comparing and mapping between two ontologies (for example, graphing neighborhoods, performing partial name matches, and searching through related document collections); (6) it can integrate and coordinate with visualization and analysis components so that knowledge can be used directly in scientific analysis; (7) it can track the development of knowledge within a community; and (8) it can serialize its own appearance so researchers collaborating remotely can share the same view into the knowledge space or can exchange their different views. This demonstration showcases some of the progress we have made recently regarding some of the above ideas, and previews our plans for the future.

USGS Activities Toward Preserving Data Collections

By Martha N. Garcia[1], Edward J. McFaul[1], Peter T. Lyttle[1], and Brenda S. Pierce[1]

[1]U.S. Geological Survey, Reston, Va.

The administration's fiscal year 2007 research and development budget priorities memorandum (issued July 8, 2005) identified Federal scientific collections—which play an important role in public health and safety, homeland security, trade and economic development, medical research, and environmental monitoring—as an area requiring special attention by the U.S. Geological Survey (USGS). As a result of the memorandum, the Interagency Working Group on Scientific Collections (IWGSC) was established in December 2005 to assess the priorities and stewardship of scientific collections and to develop a coordinated strategic plan to identify, maintain, and use Federal collections. The initial information regarding scientific collections will be gathered through a Web-based survey that will be used to develop a coordinated strategic plan to identify, maintain, and use Federal collections and to further collections research. Concurrent to the establishment of the IWGSC, section 351 of the Energy Policy Act of 2005 calls for the creation of a Preservation of Geological and Geophysical Data Program at the USGS. The first stage of this program is already underway with the formation of a Preservation Subcommittee under the Federal Advisory Committee for the National Cooperative Geologic Mapping Program. The subcommittee will help the USGS develop an implementation plan for the new preservation program, which must be submitted to Congress within one year. This plan will set priorities for several funding scenarios up to the proposed authorized funding level. A survey of agencies under the Department of the Interior and of State geological surveys is being undertaken to (1) identify at-risk geological and geophysical data holdings managed by these agencies and (2) determining the resources that are needed to rescue and make these materials available. Should the plan for data preservation be funded, the USGS will begin this effort by preserving the highest priority data collections within both the Federal and State holdings.

Gravity Reduction Spreadsheet to Calculate the Bouguer Anomaly Using Standardized Methods and Constants

By Derek I. Holom[1] and John S. Oldow[1]

[1]Department of Geological Sciences, University of Idaho, Moscow, Idaho

Current standards for the reduction of observed gravity data to a modeled Bouguer anomaly largely are unregulated and vary among geophysical textbooks, commercial software programs, and academic research spreadsheets, some of which are available for download from the Internet. Using the new standards established by the U.S. Geological Survey and the North American Gravity Database Committee, we developed a spreadsheet for the reduction of raw gravity data to the Bouguer anomaly and, with the use of terrain correction, the complete Bouguer anomaly. We view the spreadsheet as particularly useful for field data reduction and modeling where Internet access is limited or unavailable. The spreadsheet is available as a free download at the Geosciences Network (GEON) portal (http://geongrid.org/).

With the use of the Global Positioning System (GPS) for surveying the station locations and their elevations, the availability of digital terrain models, and the enhanced computational capability, gravity modeling has become a cost-effective tool in subsurface analysis of basin- to continent-scale studies. Existing gravity data for North America are archived and readily accessible via the Internet at the Pan-American Center for Earth and Environmental Studies (PACES) Web site (http://paces.geo.utep.edu/), where the gravity database provides principal facts and the Free Air, Bouguer, and complete Bouguer anomalies. The anomalies are calculated by a FORTRAN algorithm using equations and constants established by the Standards/Format Working Group of the North American Gravity Database Committee.

We developed a Microsoft Excel gravity spreadsheet in order to facilitate the adoption of the reduction and modeling standards by the research community and to provide an easy-to-use, portable gravity-correction and anomaly-computation package. Excel is a common software application used by government agencies, research institutions, and private companies and provides an easily accessible and accurate computation platform. The greatest departure of this spreadsheet from earlier gravity modeling routines is the use of ellipsoidal (rather than orthometric) height in calculations. Application of GPS in station surveys during gravity acquisition provides a direct determination of ellipsoidal height and eliminates the need to convert altitudes to orthometric heights. The ellipsoidal reference used in gravity corrections and models is Geodetic Reference System 1980 (GRS80), but the International Terrestrial Reference Frame (ITRF) and World Geodetic System 1984 (WGS84, version 2 or higher) can be substituted for the ellipsoidal datum without a significant difference in theoretical gravity values.

The gravity spreadsheet calculates the corrections for instrument drift, latitude, height above the GRS80 reference ellipsoid, atmosphere, and the Bouguer spherical cap, as well as the DC shift for multiple-day gravity surveys. The spreadsheet also supports meter-specific calibration constants for Worden analog meters, calibration tables for LaCoste-Romberg meters and digital-analog values typical of LaCoste-Romberg meters with the Aloid upgrade. Input for digital meters (Scintrex) bypasses the conversion tables, and the observed relative gravity values (in milligals, or mGals) are directly entered into the spreadsheet. Tide and terrain corrections are not calculated in the spreadsheet, but users can enter these values from other programs.

Before the new gravity standards were adopted, gravity reduction typically used orthometric heights to calculate Free Air and Bouguer-slab corrections. In the new standards, the use of ellipsoidal height removes the need to correct for the geophysical indirect effect, which compensated for the difference between the ellipsoid and the geoid. As a consequence, the Free Air correction will equal zero at the altitude of the GRS80 ellipsoid rather than at sea level.

One negative aspect of using the ellipsoidal datum in the spreadsheet is the need to convert existing gravity data that use geoidal or height above sea level to the ellipsoidal datum. To this end, we include the National Geodetic Survey (NGS) transformation (GEOID03, NGS's 2003 refined model of the geoid of the conterminous United States) to convert orthometric heights to the ellipsoidal heights used in these calculations. When combining new gravity data with existing values available from PACES, it must be recognized that the downloaded Bouguer anomaly values were calculated using orthometric heights. This can result in mismatches of as much as ±7 mGals because the separation of the geoid from the ellipsoid on the North American continent can be as great as ±20 meters. When combining gravity data downloaded from the PACES GeoNet server with new measurements, the heights of each downloaded station must be recalculated to the ellipsoid using the NGS transformation software provided with the gravity spreadsheet.

In the spreadsheet, the primary data that is required to calculate the complete Bouguer anomaly are (1) the height of the gravity station above the GRS80 reference ellipsoid, (2) the latitude of the station in WGS84 coordinates, (3) the drift- and tide-corrected, observed gravity readings tied to an absolute gravity base station, and (4) the terrain correction for the location of the gravity station. The additional calculations built into the spreadsheet include (1) the instrument drift correction, (2) the latitudinal correction, (3) an analog meter-counter conversion to relative gravity, (4) the DC shift, and (5) a conversion from local observed gravity to absolute gravity readings.

Internet Access to Spatial Data of the USGS Mineral Resources Program

By Bruce R. Johnson[1] and Peter N. Schweitzer[1]

[1]U.S. Geological Survey, Reston, Va.

Internet map services in their current form rarely provide geospatial scientific data in completely understandable ways. To date, Web map systems have focused primarily on the geographic content of the information and typically make the assumption that the user possesses complete scientific knowledge of the data through previous experience. To be usable by people who do not already have that specialized knowledge, the Web map service needs to be coupled with informative reference services (in addition to formal metadata) and must link directly to Web services that facilitate downloading the data for further scientific processing; all of this should be done possibly outside the geographic information system (GIS) environment. The U.S. Geological Survey (USGS) Mineral Resources On-line Spatial Data Web site (http://mrdata.usgs.gov) addresses these concerns by using a combination of GIS, Web, and information technologies. Scientific data are currently available to all users through three different interfaces: on interactive maps using a standard Web browser, as Web services through GIS clients, and as a selective download using an interactive database query service. All of these interfaces use consistent terminology and explanations.

We will demonstrate the three currently available user interfaces along with an experimental service that is compliant with the Open GIS Consortium and will show overlays of several data layers coming from geographically separate servers. All data layers in the experimental service are presented within a consistent geoscience data model (Geoscience Markup Language, or GeoSciML). This model is being developed by the Data Model Task Group, which is an international collaboration under the auspices of the International Union of the Geological Sciences' (IUGS's) Commission for the Management and Application of Geoscience Information (CGI).

Online Access of Publicly Available U.S. Geological Survey Geophysical Data

By Robert Kucks[1]

[1]U.S. Geological Survey, Denver, Colo.

The U.S. Geological Survey (USGS), through its many regional centers, has released to the public a varied and extensive amount of geophysical data. Most data from the late 1990s through the present have been available in reports published digitally on the Internet by the USGS. Datasets from these reports include potential fields (magnetics and gravity), electrical methods (electromagnetic and resistivity), natural fields (magnetotelluric and audiomagnetotelluric), radar, and gamma-ray (radiometrics). A Web site has been developed by the Mineral Resources Program's Crustal Imaging and Characterization Team to allow simplified access to all these datasets. For magnetic data, there are two sites that contain digitally obtained and digitally converted map (digitized) data for the conterminous United States. Linked access is also available to gravity station data at a site hosted by the University of Texas—El Paso's Pan-American Center for Earth and Environmental Studies (PACES) in cooperation with the National Geospatial-Intelligence Agency, the National Oceanographic and Atmospheric Administration, USGS, industry, and academic colleagues. The formats of these geophysical datasets are generally in ASCII or text forms that permit common accessibility to the general public.

Discovery and Semantic Integration of Geologic Data

By Kai Lin[1] and A. Krishna Sinha[2]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[2]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Va.

The frontier of cyberinfrastructure research in the geosciences has brought geoscientists opportunities that range from implementing new and better ways to access data from distributed systems (including issues of data interoperability), to accessing tools for computations and visualization, to bringing legacy data into the digital world. None, however, is as significant for a geoscientist as the ability to integrate data from a wide variety of disciplines such as petrology, stratigraphy, structure, geophysics, and hydrology. We present the results of our ongoing research into providing the infrastructure that is required for the discovery and integration of certain types of data through the use of ontologies.

Although there are many definitions of ontologies, we use the term to explicitly represent relationships in information contained in heterogeneous and multidisciplinary datasets. We have developed seven high-level packages of ontologies in Web Ontology Language (OWL): Planetary Materials, Planetary Phenomenon, Planetary Structures, Planetary Location, Physical Properties, Geo-Image, and Geologic Time. These packages provide the framework for capturing concepts and relations contained both within and between datasets. For example, registering data to ontology for rock geochemistry, rock texture, rock mode, and rock location allows for schema-independent integration between chemical, modal, and textural datasets. This integration requires that data providers register the data to ontologies because registration allows users to query the datasets through the ontologies.

The registration procedure, utilizing ontologies, resolves the syntax, structure, and semantic heterogeneities of datasets. The mediator (a software package to send queries to distributed data resources) allows a user to issue an ontologic query that is independent of any schema; the mediator rewrites the ontologic query into a physical query accepted by different data resources (for example, Oracle, DB2, SQL Server, MySQL, as well as MS Excel files converted to databases, such as PostgreSQL) and wraps the results into a uniform format. This framework for data discovery and semantic integration requires geoscientists to formalize the domain knowledge into ontologies, register data to ontologies, and allow computer scientists to develop mediators whose function is to help provide a more integrative view of the dynamics of the Earth or other planets.

A Rapid Registration Tool—Making Legacy Map Data Geospatial

By Ben Logan[1] and A. Krishna Sinha[1]

[1]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Va.

With the emergence of digital processing of information in earth science, it is important to provide access to many types of nondigital legacy data. Such data include information in tables contained in publications as well as maps and images that are not amenable to geospatial queries. For example, although it is possible to transfer scanned images of maps to a Web environment, the information contained in the maps—such as sample locations—has to be visually estimated. Because sample locations provide the geospatial context for the geologic analyses that were conducted on these samples, it is useful to provide a tool to make the locations compatible with modern geographic information system (GIS) search capabilities.

We have developed a technique for rapid geospatial registration of sample locations or field observation sites for legacy data. The Rapid Registration Tool (RRT) provides a simple interface for locating and labeling sample sites from scanned images. Our software provides for scanned images to be registered to U.S. Geological Survey (USGS) 7.5-minute quadrangles, which are available as digital raster graphics (DRGs) for all the States of the conterminous United States, so that points from a scanned image can be projected to the DRG. The resulting locations (in latitude and longitude) of individual points can be inserted into existing data tables, thus providing the geospatial context required for GIS-based queries. The output from this process will be a shapefile in the same projection (Universal Transverse Mercator projection and corresponding zone) as the underlying DRGs, which allows for further use in ArcGIS, ArcView, or freely downloadable software, such as ArcExplorer.

The principle for registering these data is an affine transformation, which moves, scales, rotates, and skews data evenly. DRGs for the conterminous United States are available on CDs from the Geosciences Network (GEON) project (www.geongrid.org), which is supported by the National Science Foundation. The user interface provides query tools for locating the desired DRGs, which then may be downloaded from the CD. Next, the user uploads a scanned image of the legacy map, which is automatically placed in one window while the corresponding DRGs are loaded into an adjacent window. A pointer tool allows the user to make links between features on the unregistered scanned image and the same features on the DRGs. Multiple pairs of corresponding features are necessary for improved accuracy of located field points. Zoom and pan tools are provided for precise selections of features. A table format provides an easy interface for monitoring the cumulative error associated with the registration of known features as well as a simple form for deleting points with large residual errors. The registration of features is followed by matching points from the legacy map to a data file that contains the sample number, longitude, and latitude. For increased ease of use, the DRG window—after registration—“follows” any zooming or panning the user performs in the first window. In this manner, the user will always be certain about where the projected points will land on the DRGs.

The usefulness of this tool is enhanced by its simple interface, which provides any user who is not familiar with GIS the ability to enter data into the ever-growing database for digital spatial analysis. The contribution this tool will make to the field of geology in general will be an increase in the number of available georeferenced sampling and analysis points that are currently only aggregated or subdivided; future researchers and educators will be able to use this information with little replication of effort.

Strike Dependence of Variable Scarp Height on the Emigrant Peak Fault, Western Great Basin—Limitations of Digital Elevation Models and Digital Orthophoto Quadrangle Analysis

By Mark M. Nell[1] and John S. Oldow[1]

[1]Department of Geological Sciences, University of Idaho, Moscow, Idaho

The Emigrant Peak fault system, located along the western flank of the Silver Peak Range in west-central Nevada, is part of an active displacement transfer system linking the right-lateral Furnace Creek-Fish Lake Valley fault system to northwest-trending transcurrent faults of the central Walker Lane. The fault is part of a northerly striking system of down-to-the-west high-angle faults that parallel the western range front of the Silver Peak Range for tens of kilometers before swinging northeast and cutting across the mountains. The Emigrant Peak fault is the youngest splay of the Pleistocene to Holocene fault system that records a history of fault activity that progressively stepped westward from the range front into adjoining Fish Lake Valley through time. The Emigrant Peak fault is the active strand of the system and cuts late Pleistocene to late Holocene alluvial fan deposits, with the youngest alluvial surfaces exposed in the study area along the southern reaches of the structure. In the footwall, alluvial fans are low-relief surfaces that experienced only localized dissection by younger channels. In the hanging wall, the downthrown alluvial surfaces are locally buried, particularly near the mouths of younger canyons, and exhibit variations in surface elevation of up to 10 meters (m) over horizontal distances of 200 to 500 m. Over a distance of 5 kilometers (km), the fault has a curvilinear trace that, from south to north, varies from northwest around a broad bend (convex west) to northeast and through another bend (concave west) to north. The fault trace is characterized by a pronounced scarp that ranges from 6 m to 30 m in height, with the highest elevations developed along north- to northeast-trending fault segments, and the lowest elevations found along the northwest-trending segment. The fault scarp morphology is well preserved and dips about 60° on north- to northeast-trending segments and decreases in dip to 35° along northwest-trending segments. The fault scarp morphology is planar for most of the trace; however, in the northern parts of the fault, it takes on a faceted morphology with a steep lower slope (15 to 25 m high) and a higher, older scarp (10 m high) with a degraded morphology and a dip of 30° to 40°.

Using directly measured heights at a few locations and elevations from a series of profiles taken from 10-m-resolution digital elevation models (DEMs), we document that the scarp height varies systematically with strike. Using the normal to the maximum scarp height as a reference, the height decreases with the azimuthal cosine of the local scarp-normal orientation around the curved fault trace. The cosine variation in scarp height assumes constant dip of the fault at all azimuths. Small deviations from the ideal cosine relationship observed along the lower northwest-trending scarps are consistent with a sine function variation in height attributed to differences in vertical and horizontal components of slip resolved on fault surfaces with decreased dip.

Comparisons of direct scarp-height measurements with those determined from DEMs indicate a vertical variance of about ±3 m. This level of elevation uncertainty is consistent with our assessment of elevation noise found in the 10-m-resolution DEM coverage. For estimation of absolute elevation, 10-m-resolution DEMs based on 15-minute topographic maps have an uncertainty of ±7.5 m but typically are internally consistent at a range reported by the U.S. Geological Survey (USGS) as ±1 to 2 m. Our analysis of surface-elevation resolution indicates that vertical uncertainty varies as a function of slope. For slopes of greater than 50° that cover two or more 10-m-resolution DEM cells, uncertainties are small and consistent with USGS estimates. On slopes less of less than 40°, we recognize a terracing effect on the surface with maximum amplitude of about ±3.5 m. This undulation is a well-recognized artifact of the cubic spline interpolation algorithm used in the production of digital elevation models. The vertical uncertainty is suppressed on slopes between 20° to 50° and decreases progressively from a maximum of ±3.5 m to ±1 to 2 m.

The scarp height and surface dip of the Emigrant Peak fault are sufficient to allow resolution of faceted morphology. In the northern part of the fault, the upper facet has an amplitude of about 10 m and is readily picked out from DEM-generated scarp profiles. We estimate that a faceted surface with a height of about 4 to 5 m or greater is detectable with a 10-m-resolution DEM. Measurement of the actual slope of fault scarps is not possible with 10-m-resolution DEM data and dips of surfaces with differential elevations of 12.2 m within contiguous cells are underestimated by up to ~40°, but the underestimation decreases to 10° for features reaching differential heights of 75 m.

DEM analysis is greatly strengthened when combined with digital orthophoto quadrangle (DOQ) coverage. The 1.0-m-resolution of the pixels that make up the geospatially rectified aerial photography greatly enhances the ability to map subtle variations in surface morphology at scales below the resolution of 10-m-resolution DEM coverage. DOQ coverage revealed that the younger alluvial fans locally buried the hanging-wall expression of the surface offset by the fault scarp to depths of at least 10 m. Thus, measurement uncertainty is not restricted to data resolution but also arises from physical processes and emphasizes the need to combine digital tools in order to improve analysis precision and accuracy.

Resource and Environmental Dynamics—The Role of GIS in Monitoring, Analysis, and Management in a Case Study from Nigeria, Sub-Sahara Africa

By E.G. Olagunju[1]

[1]Department of Computer and Information Science, Linkoping University, Linkoping, Sweden

This paper is a description of the work done on the role of geographic information systems (GIS) in the management of resources and environmental dynamics. The context of study is Nigeria, which is a microcosm of sub-Saharan Africa. It is noteworthy that this region of the world is richly endowed with natural resources (water- and land-based) and environmental variables that are of global interest; however, it is a disadvantaged region because of poor analytical, monitoring, and management technology of these resources and the environment. Poverty, environmental pollution (water, oil, and gas), biodiversity degradation (change in climatic conditions and global warming), poor livelihood, security concerns, and disasters characterize this third-world region. The impact of these indices of underdevelopment on this region jeopardizes the ability of the population to benefit from the region's resources.

This demonstration focuses on the role of GIS in the effective analysis of the natural resources (water, land, and so on) and their deposition or distribution in Nigeria, where natural resources are inequitably distributed. The important points include the following: (1) the interface between GIS and resource monitoring, especially during exploration and utility of the resources, (2) the role of GIS in monitoring key environmental variables that are essential for environmental preservation and management, (3) specific relevant applications in agriculture and water-resources management, and (4) enumerating the current GIS capacity of a region as a way of interpreting the current trends and giving a basis for the future. The recommended use of GIS technology in research, training, and production of empirical data for local, regional, and global policy formulation also will be discussed.

Noesis—An Ontology-Based Semantic Search Tool and Resource Aggregator

By Rahul Ramachandran[1], Sunil Movva[1], Phani Cherukuri[1], and Sara Graves[1]

[1]Information Technology and System Center, University of Alabama—Huntsville, Huntsville, Ala.

Current search tools typically use text string comparisons to find resources matching a user's needs. For example, searches on science data catalogs are typically based on a controlled vocabulary or a specialized set of search keywords. Smarter search tools can be developed that use ontologies to provide semantic capabilities. An ontology encodes concepts and the relationships among them. From a machine learning or artificial intelligence perspective, it is viewed as a formal, explicit specification of a shared conceptualization. It captures consensual domain knowledge of concepts and constraints of their use in a machine-understandable fashion. As part of the Linked Environments for Atmospheric Discovery (LEAD) project, we have developed an ontology-based semantic search tool for earth science. The ontology used by this tool currently focuses on meteorology. The tool supports both ontology browsing and ontology-based semantic search capabilities. While browsing, users are able to navigate the meteorology ontology from different points of access. Thus, a user can start at a general topic and navigate to specific topics by selecting the related concepts of interest. The ontology-based semantic search provides several inference capabilities, such as equivalence, inversion, generalization, and specialization. These capabilities support an expanded search vocabulary in addition to the list of controlled keywords. The tool uses these inference capabilities to serve as a one-stop shop to collate resources, such as Web pages, datasets, and publications relevant to meteorology. When a user enters a search term, our tool first uses the ontology to find synonyms and other related concepts, and then uses the results from this inference step to query other resources—such as search engines like Google and Yahoo, educational databases like DLESE, data catalogs, or a scientific journal database—to find all the relevant resources.

The Paleontological-Stratigraphic Interval Construction and Analysis Tool

By Joshua A. Reed[1], Cinzia Cervato[1], and Douglas Fils[1]

[1]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

The Paleontological-Stratigraphic Interval Construction and Analysis Tool (PSICAT) is a Java-based graphical editing tool for creating and viewing stratigraphic column diagrams from drill cores and outcrops. The PSICAT is designed around a modular, plug-in-based architecture that allows the tool to be extended to support a wide variety of functions, tasks, and geoscientific communities. Through the use of different combinations of plug-ins, the tool customizes itself to a specific task (such as drawing stratigraphic columns) and captures data digitally as the user draws and edits the diagrams. The data are captured in open, Extensible Markup Language (XML) formats. CHRONOS (http://www.chronos.org) will be providing a central database for the data and diagrams. Users of the central database can make their data and diagrams accessible to anyone, anywhere, at any time. Another goal of PSICAT is to integrate with some of the tools accessible via the CHRONOS portal, such as the CoreWall Suite.

PSICAT is currently being developed for use by the Antarctic Geological Drilling (ANDRILL) project (http://www.andrill.org/) on their upcoming drilling expeditions in Antarctica, but a general community version will also be available. The PSICAT will allow unprecedented communication between Antarctica-based scientists and shore-based scientists, potentially allowing shore-based scientists to interact in almost real time with on-ice operations and data collection. A planned generic version will have wider applications for field work and core logging in industry and academia.

Online Databases and Paleomaps

By Allister Rees[1] and Javier Espinoza[1]

[1]Department of Geosciences, University of Arizona, Tucson, Ariz.

We have assembled five global-scale databases comprising sedimentary rocks, dinosaurs, and fossil plants, which are Web-accessible via the Geosciences Network (GEON) portal and at the University of Arizona (http://www.geo.arizona.edu/~rees/). The databases contain about 175,000 records worldwide; sedimentary rock data span primarily the Permian to Holocene; dinosaur data, the Triassic through Cretaceous; and plants, Permian through Jurassic.

The databases can be searched individually or in any combination with the others, via text or interactive maps. They can also be searched by any combination of age, location, or content as well as reference source (author, year, title). Search results can be downloaded as text or viewed on present-day maps and paleomaps.

A primary goal is to facilitate "cross-fertilization" between disciplines by enabling easy and seamless retrieval of information about geographic patterns of climate-sensitive sediments, animals, and plants through time. In this manner, fossils can be studied in the context of geography and climate, paleogeographic reconstructions can be refined, and global paleoclimate models can be compared with a wide range of proxy data. These databases represent a first step in this direction and form part of the broader GEON project.

Sedimentary rock data were compiled primarily by Alfred Ziegler (University of Chicago) and other members of the Paleogeographic Atlas Project (PGAP), with additional oil source-rock data provided by Judith Parrish (University of Idaho). Dinosaur distributions are from "The Dinosauria" (Weishampel and others, 2004). The fossil plant data are from the Paleobiology Database, which is operated out of the University of California at Santa Barbara (http://www.paleodb.org/cgi-bin/bridge.pl). Paleomaps and plate rotation programs are courtesy of Christopher Scotese (PALEOMAP Project).

Reference Cited

Weishampel, D.B., Dodson, Peter, and Osmolska, Halszka, eds., 2004 [2nd. ed.], The Dinosauria: Los Angeles, University of California Press, 861 p.

Analyzing EarthScope USArray Data Using GEON Cyberinfrastructure Network Resources

By Dogan Seber[1], Tim Kaiser[1], Choonhan Youn[1], and Cindy Santini[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

EarthScope is a nationwide, multi-agency project, funded by the National Science Foundation, that applies "observational, analytical and telecommunications technologies to investigate the structure and evolution of the North American continent and the physical processes controlling earthquakes and volcanic eruptions" (http://www.earthscope.org/overview/index.php). As EarthScope continues to install broadband seismic stations across the western United States, a significant challenge awaits the user community in processing voluminous data to be collected by USArray (a continent-wide seismic observatory that is a component of EarthScope). To alleviate some of the difficulties in accessing and processing the collected data, we have built a geoinformatics platform (the synthetic seismogram tool, or SYNSEIS) within the Geosciences Network (GEON) cyberinfrastructure. The platform not only provides access to seismic datasets, but also helps researchers to model and analyze collected earthquake waveforms. Users could calculate synthetic seismograms of any of the regional earthquake recordings and study the lithospheric structure. Potential topics of study include the source depth estimations, source dynamics, propagation anomalies, Q structure, and seismic velocities. In this study, we have analyzed recordings of a limited number of southern California earthquakes recorded by USArray and modeled the waveforms to test the tool and to provide a better understanding of seismic propagation in the region. We will demonstrate the tool and outline our findings at this presentation.

Architecturally, SYNSEIS is based on a Web services model and utilizes the E3D finite difference code developed by Shawn C. Larsen and others at the Lawrence Livermore National Laboratory to compute the synthetics. The user interacts with the SYNSEIS user interface through a Web browser, which accesses a central user interface server that contains the collection of Web service clients. Using a built-in three-dimensional (3D) Earth structure, SYNSEIS allows (1) building two-dimensional (2D) and 3D models, and (2) computing seismograms using distributed resources available within GEONgrid as well as the national TeraGrid system. In order to run the applications, the user first selects model parameters and computational resources and creates input data parameters, which are kept as Extensible Markup Language (XML) objects. These parameters are then sent to the computational platforms, and jobs are executed automatically without additional user interaction. Because this is built as a GEON resource, the system takes advantage of other GEON services as well. For example, each job is archived in the myGEON area of the GEON portal. Users could share the inputs and (or) outputs of their models with others and rerun any job from the myGEON area.

NASA's Global Change Master Directory (GCMD)—Managing Metadata for Discovering and Accessing Earth Science Data and Services in Support of the National Cyberinfrastructure

By Tyler Stevens[1]

[1]Global Change Master Directory, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Md.

The National Aeronautics and Space Administration's (NASA's) Global Change Master Directory (GCM) contributes to the cyberinfrastructure efforts by enabling discovery and access to earth science datasets and services. Through a robust and flexible international metadata format (compatible with Federal Geographic Data Committee (FGDC) standards; the International Organization for Standardization's ISO 19115, which is the schema required for describing geographic information and services; and the Dublin Core Metadata Initiative), the GCMD helps determine if a dataset meets the user's needs. The discovery of datasets is made available through intuitive options, such as controlled earth science keywords with optional query refinements, a free-text search, and a geospatial map search. Another option for searching is through portals. Portals are subsets of metadata content in a customized interface. A set of well-defined earth science and service keywords has been used in numerous semantic Web applications and ontology development to promote interoperability across the earth science discipline. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is also supported to promote interoperable metadata harvesting. Online authoring tools allow users to create and submit metadata to the GCMD. This demonstration will show how to use the GCMD interface to discover, access, and manage metadata related to earth science datasets and services.

Exploratory Geovisualization in Improvise

By Chris Weaver[1]

[1]Department of Geography and the GeoVISTA (Geographic Visualization Science, Technology, and Applications) Center, The Pennsylvania State University, University Park, Pa.

The development of software tools for rapid visual analysis is a recent research focus in geographic visualization. The goal is to allow users to explore information that contains geospatial, temporal, and abstract components in a flexible, integrated, and interactive environment that requires minimal training to use. "Improvise" software is an exploratory visualization environment, written in Java and freely available as Web-based open-source software. In Improvise, users build and browse multiple, coordinated views of their data. What makes Improvise special is that it offers precise control over how navigation and selection affects the appearance of space, time, and abstract attributes in and between multilayer maps, scatter plots, parallel coordinate plots, tables, spreadsheets, and other views. Building and browsing activities are integrated in a single live-user interface that lets users alter views quickly and incrementally. Entire visualizations can be rapidly modified and extended to develop hypotheses and exploit discoveries during ongoing exploration and analysis. Improvise visualizations are saved as regular Extensible Markup Language (XML) text documents in a self-contained, platform-independent format for easy sharing and dissemination of results.

This poster session and demonstration will describe the process of interactive visual exploration and analysis using multiple geovisualization applications currently being developed in Improvise as a part of ongoing research at the GeoVISTA Center at The Pennsylvania State University. Historic hotel-guest visitation patterns, ham radio communications, Michigan hydrographic information, county-level election results, and the U.S. Census are a few of the datasets under current investigation. These efforts involve extending the single-user multiple-view approach used in Improvise to build a heterogeneous collaborative geovisualization architecture for time-critical applications, such as emergency management and intelligence analysis.

Application of Text Mining in Developing Standardized Descriptions of Taxa in Paleontology—A Framework

By Wen-Bin Yu[1], Francisca E. Oboh-Ikuenobe[2], and Bih-Ru Lea[3]

[1]Department of Information Science and Technology and Department of Computer Science, University of Missouri–Rolla, Rolla, Mo.

[2]Department of Geological Science and Engineering, University of Missouri–Rolla, Rolla, Mo.

[3]Department of Business Administration, University of Missouri–Rolla, Rolla, Mo.

Like other disciplines of science, the the discovery of new information and the modification of existing knowledge enables advancements in the field of paleontology. The process of discovery of new information generates large volumes of data that can be overwhelming if not properly stored and (or) utilized. For example, the Treatise on Invertebrate Paleontology created by Professor Raymond C. Moore at University of Kansas blazed the trail for similar works that came later. Many paleontological volumes provide information on fossil specimens that have been formally named. In palynology, problems can arise with palynomorph classifications and interpretations because of the subjective nature of classifications due to human judgments and different levels of training. As a result, the same palynomorph can be interpreted or classified differently, resulting in junior synonyms and amended descriptions that can potentially confuse students and new researchers. It is important to provide a framework to compose a standardized description of each taxon using diverse observations from various taxonomists.

The main objective of this study is to propose a framework that uses text mining techniques in developing a taxon description recommendation system. Text mining can apply intelligent methods and algorithms to extract or mine knowledge and meaningful data patterns from a large amount of unstructured texts or documents for decisionmaking; therefore, it is expected that common characteristics and features from interpretations done by different scholars can be captured and used for clustering and description to minimize the issue of subjective human judgment.

The proposed framework will be illustrated using a sample database and a tutorial example. This study will provide insights on (1) how text mining can be used to develop a descriptive model, and (2) how descriptive terms generated during the text mining process can be used to provide a basic set for a standard lexicon to develop a standardized taxon description recommendation. Furthermore, advantages and drawbacks of the proposed framework will be discussed, and future research directions will be proposed.

Ontology-Driven Geospatial Service Composition Using a Multiagent Framework

By Genong Yu[1], Liping Di[1], Wenli Yang[1], Yue Peng[1], Peisheng Zhao[1], and Aijun Chen[1]

[1]Center for Spatial Information Science and Systems (CSISS), George Mason University, Greenbelt, Md.

Solving geospatial problems in Web service environments normally requires invocation of a chain of geospatial services. The process of creating the chain is called geospatial Web service composition. Most Web-service-composition approaches rely heavily on a centralized directory. The discovery of agents or services depends on the completeness of the directory. The network traffic load for the centralized directory service can grow tremendously with increasing requests and may eventually be out of capacity. Ideally, the discovery and composition can be restricted to access local information. To enable the scalability in a distributed environment, the composing process and interactions should be limited to local information without heavy demand for global information. This study strives to explore the possibility of using a multiagent system framework to tackle this problem. A multiagent system is scalable because it can handle from several to hundreds or thousands of agents in a consistent approach. Through a multiagent system, it is possible to implement a discovery mechanism that relies solely on local information. For example, agents may know their neighbors only by keeping a short list of known agents, and messages can be conveyed through a mesh-like network topology. Such scalability is achieved at a cost of some delay due to interactions among many agents. To demonstrate the feasibility of such an agent system, an example scenario of landslide susceptibility assessment is implemented under the Java Agent Development Framework (JADE). Each individual agent is autonomous with capabilities as a receiver and a requester. Capabilities or services provided by each agent are conveyed with ontology internally managed and matched through Java objects and externally through the standardized Foundation for Intelligent Physical Agents Semantic Language (FIPA-SL). Through such a framework, each agent is only responsible for its capability and a short list of "acquaintances." Querying to any agent could lead to the correct answer or a chain of services given that there is a large repository of data-serving and data-processing agents. The request is either answered or relayed through a standardized message-forwarding mechanism. The study shows the scalability and efficiency of such a framework. Further study will examine optimization of agent integration, semantic alignment, and interoperation of agent and Web service.

Path Planning for Automatic Chaining of Geospatial Web Services

By Peng Yue[1,2], Liping Di[1], Wenli Yang[1], Genong Yu[1], and Peisheng Zhao[1]

[1]Center for Spatial Information Science and Systems (CSISS), George Mason University, Greenbelt, Md.

[2]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

Semantic Web technologies provide a promising prospect for automatic discovery and chaining of geospatial Web services. A semantic-enriched geospatial Web service, represented using semantic Web service technologies, is called a geospatial semantic Web service. A graph can be formulated through information from multiple geospatial semantic Web services. Nodes in the graph represent services and connectivity or edge weight is determined by the semantic matching of input and output of the services. Path planning is introduced in the graph to enable the automatic chaining of geospatial Web services. The method is helpful for answering particular geospatially related "what if" questions in a Web service environment. In path planning, the decision and execution phases are usually differentiated. Thus, the chaining process in path planning includes three phases: (1) path modeling, which involves constructing paths in the graph as plans between start and end nodes (for example, the "if" condition and the "what" part in the question can be conceptualized as start and end node, respectively); (2) plan instantiation, where the plan is instantiated into a concrete workflow. An individual service is selected at each node automatically based on the service's advertised semantic capability and the related quality-of-service (QoS) information; and (3) workflow execution, where the chaining result or workflow is executed in the workflow engine. These three phases interact with each other. The failure of execution at the posterior phase can notify the upper phase for an alternative selection. This recursive process continues until a plan is applicable. A prototypical system is implemented to demonstrate the concept and approach of this research. The OWL-based Web service ontology (OWL-S) is adopted for representing geospatial semantic Web service. The service profiles in OWL-S descriptions provide the nodes and edges in the graph. Multiple paths as alternative plans are generated using a K-shortest path algorithm. The individual services with the same service profile can be selected based on QoS information. A unified interface for QoS providers is supported so that different QoS criteria can be plugged in. The concrete workflow can be converted to Business Process Execution Language (BPEL) for Web Services to facilitate the use of the BPEL engine. A use case on landslide susceptibility assessment has been employed in this system. Six types of Web services—including Open GIS Consortium, Inc.'s (OGC's) Web Coverage Service (WCS) and Web Image Classification Service (WICS)—are involved in the final service chain. The study shows the applicability and efficiency of this solution. Further study will focus on the refinement of a plan to support more specified domain logics or rules.

Posters

Only the abstracts from each poster session are included below. To view the posters, go to the Geoinformatics 2006 meeting Web site at http://www.geongrid.org/geoinformatics2006/.

The Role of Microstructure Development on Crystal Size Distribution and on Recovering Crystal Size Distribution From Thin Slices—A Modeling Approach

By Roddy Amenta[1], Sarah Roberts[1], and Paxton Wertz[1]

[1]James Madison University, Harrisonburg, Va.

Crystallization experiments using computers are most useful when direct experiments with melts are intractable, such as in analysis of microstructure development. For this reason, computational modeling was employed to simulate the nucleation and growth of crystals competing for space to form polycrystalline microstructures. The theory that quantitatively links the crystallization kinetics with the resulting crystal size distribution (CSD) is well established, but it is based on the premise that crystals grow freely with euhedral shapes surrounded by large amounts of mother liquid or magma and that they maintain fixed aspect ratios and shapes. The extension of this theory to crystallization that produces microstructures is less certain because the above conditions do not hold, and crystal shapes and sizes are modified by their grain-to-grain impingements. We modeled crystal growth by adding unit cells to their lattices and modeled linear CSDs by coupling constant crystal growth rates with exponentially increasing nucleation rates. Thus, crystallization advanced to fill more than 90 percent of the chamber with three-dimensional microstructures.

Seven crystallization models were run, and each one produced microstructures composed of unique crystal shapes, such as prisms (1:1:5 and 1:1:3), plates (1:5:5 and 1:3:3), and cuboids (1:2:5, 1:3:5, and 1:4:5). Crystal sizes and shapes were measured from digital thin slices through the microstructures as would be done from rock thin sections, and stereological corrections were used on these data to attempt to recover the CSDs. The true-size CSDs were calculated from the number of unit cells contained in each crystal. The recovered CSDs were then compared to their known true-size CSDs, and the true-size CSDs were compared to their respective predicted or ideal CSDs. An unexpected discovery is that all the modeled microstructures showed a very close correspondence between the true and ideal CSDs, which indicates that the microstructure development did not significantly affect the CSD information. The recovered CSDs were linear for the prisms, slightly curvilinear for the plates, and modestly curvilinear for the cuboids. The recovered CSD slopes and intercepts corresponded closely to the true-size CSDs in the prisms and plates, but those for cuboids showed mixed results. We conclude from these results that rocks with non-linear CSDs should be interpreted cautiously as indicators of nonlinear crystallization histories and that petrographic examination should have precedence in such interpretations.

Ontology of Experimental and Simulated Fault Gouge

By Hassan A. Babaie[1] and Jafar Hadizadeh[2]

[1]Department of Geosciences, Georgia State University, Atlanta, Ga.

[2]Department of Geography and Geosciences, University of Louisville, Louisville, Ky.

We have designed ontologies related to an investigation of changes in gouge porosity in small displacement faults via two dimensional (2D) computer simulations and experimental deformation. These ontologies, which are built based on the state of the art domain models, will facilitate efficient online interchange and reuse of knowledge that results from research of physical processes and materials involved in the deformation. The groundwork consists of ontologies for brittle rock deformation that help the understanding of deformation along natural and artificial brittle fault zones. These ontologies are formal, controlled vocabularies that reflect domain theories, and are developed for shared use within and across domains. The vocabularies include terms that represent real world individual entities, such as a particular fault, mylonite, or gouge zone, and formal relations such as subsumption, parthood, participation, location, and agency between these individuals. Instances (individuals) of these material and immaterial entities have all their spatial parts present at all times of their existence, and occupy same or different spaces at different times. These entities, which are sampled by geologists at specific time instants and locations, may change their properties and qualities without losing their identity. For example, a fault may change its length, or a gouge may change its density or porosity, or become hot. Some entities evolve, over time, into new entities with new identity, through transformation. For example, a particular granite may deform into a mylonite, or a 2D surface may bend into a fold. Process entities, such as deformation, cataclasis, and slip localization involve material entities as agents and participants. For example, water is an agent during hydrolytic weakening and hydraulic fracturing, and rocks participate in fracturing or alteration. Process entities have no spatial parts, compared to the material entities that do not have temporal parts. The temporal parts of the process entities unfold over time in several phases. The spatial location of material entities that participate in particular processes define the location where these processes occur. Processes may destroy, or create material entities, over different times and under different conditions.

Our ontologies are designed with a modular architecture, allowing integration and a more realistic depiction of the interaction of individual entities during deformation in simulated, experimental, or natural gouge. The ontologies include the following classes and their object properties (some of which are given in parentheses): fault, gouge (porosity), region, particle (shape, number, contact, sliding surface area), particle size (distribution, fractal), packing (density), and gouge texture. The process classes include the following: shear displacement, shear localization, comminution, weakening, friction, cataclastic flow, and fracture. We present the vocabulary (classes and relations) of each of these non-overlapping ontologies from the processual and material perspectives.

From Service Replica to Service Grid—Lessons Learned From Building LAITS OGC-Compliant GeoGrid

By Yuqi Bai[1], Liping Di[1], Aijun Chen[1], Yang Liu[1], and Yaxing Wei[1]

[1]Center for Spatial Information Science and Systems (CSISS), George Mason University, Greenbelt, Md.

The emerging service-oriented computing (SOC) paradigm brings a new model of software reuse by means of standard Web services. It enables the building of composite services from multiple Web service instances. Efficient service composition is extremely challenging due to the dynamic characteristic of the network and the unpredictable quality of service (QoS) of participated services. Many related studies have been focused on building appropriate QoS models for Web services and proposing suitable computing network models for particular application needs. While most of these studies assume that Web service instances are closely connected with their hosts, we believe that service instances may be dynamically deployed on new machines other than on their hosts and, in some cases, at the run time. To investigate this new computing mechanism, we have developed a proof-of-concept system, the Laboratory for Advanced Information Technology and Standards (LAITS) OGC-Compliant GeoGrid, using the Globus toolkit.

This poster presents the details of design and implementation of this system, with a focus on how service replicas are managed. A case study of application of this computing mechanism on the composition of Grid services is also introduced. The performance metrics show that even though there is some overhead owing to the dynamic service deployment and configuration, this mechanism may enhance the overall performance of the composite Grid service, from composition to execution, by deploying participated services on new host nodes to provide more candidate composition plans. At the end of this poster, we discuss the constraints applied to implement this new mechanism in a production environment and the potential application of the mechanism in building a service Grid or the fully automated service-computing environment.

From Object to Process Ontology

By Calvin G. Barnes[1]

[1]Department of Geosciences, Texas Tech University, Lubbock, Tex.

The recent National Aeronautics and Space Administration (NASA)-supported Semantically Enabled Science Data Integration (SESDI) workshop outlined the first ontologic platform for classification of volcanoes, volcanic activity, and eruption phenomena. One of the fundamental issues raised at this workshop was the relationship between objects (for example, volcanic products) identified within an existing ontology and processes that provide cause-and-effect linkages between volcanic activity and volcanic products. A major goal of earth-science-related ontologic research is to link processes to objects in a way that will ultimately provide predictive capabilities. In the case of volcanology, this ability could provide for information technology-based integration of geologic data, historical eruption records, real-time monitoring information, and petrologically based process models to provide decision-makers with the input they need during volcanic crises.

For domain experts, investigation of geologic processes typically involves data-driven inference, specifically, the use of detailed datasets that involve objects, time, and space to derive testable genetic hypotheses (“reverse” models). In igneous petrology, reverse process models require the ability to use field, geochemical, and mineralogical data to infer magma source(s), storage site(s), and causes for compositional diversity. This is generally done by thermodynamic calculations to determine pressure (P) and temperature (T) and by matching patterns of compositional arrays to infer differentiation/assimilation processes. By a combination of logic and trial-and-error comparison, researchers eliminate as many genetic scenarios as possible. If coded for hands-off computation, this approach borders on that of artificial intelligence.

In contrast, predictive models require the ability to visualize and calculate a process from its inception to completion (“forward” models). Forward modeling of magmatic processes is similar to workflow calculations. In the simplest approach, magmatism is a bottom-to-top phenomenon, from a mantle heat source to final pluton emplacement or eruption; however, although adding complexity to forward models is relatively straightforward, the usefulness of such complex models may suffer owing to a lack of general applicability in complex volcanic environments.

If we use petrology and volcanology as examples, then the challenges to developers of geologic process ontologies include the following:

- 1. Is it possible for a single form of process ontology to accommodate reverse and forward models?
- 2. How do objects interact with process? Are interactions purely semantic or are they numerical?

- 3. Can a process ontology that involves objects be written in such a way as to be nondirectional (that is, either forward or reverse)?
- 4. Can ontology languages such as Web Ontology Language (OWL) accommodate numerical calculations? If so, what is the level of complexity? If not, can workflows be written to handle reverse, inference-based logic?
- 5. How can experimental data and (or) theoretical computations be integrated into process ontologies?

GEON—Cyberinfrastructure in Support of a Distributed Geosciences Community

By Chaitan Baru[1], Karan Bhatia[1], and Ashraf Memon[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

The Geosciences Network (GEON) was funded as an information technology (IT) research project to investigate the cyberinfrastructure needs of the earth sciences community. The details of the distributed system architecture and related system policy issues have crystallized over the duration of the project as a result of close collaborations between geoscientists and computer science and IT researchers.

In this poster, we present the design and update the status of the GEON cyberinfrastructure, which is needed to support a distributed geosciences community. Our goal has been to define a set of high-level geosciences services and a variety of end-user tools for accessing these services. These services may themselves evolve along with the needs and requirements of the geosciences user community. The overall system design is based on a careful balance between centralized and distributed control and services.

We present cyberinfrastructure solutions developed in GEON for the various systems challenges, which include the following:

Security and Access Control. The Grid Account Management Architecture (GAMA) has been developed to improve the management of Grid user accounts. Custom Java Database Connectivity (JDBC) drivers provide restricted access to datasets. Access to the portal and portlets is based on the user's role. Web services with Grid Security Infrastructure (GSI) authentication support authenticated access to system services.

Systems Management. The widely used Rocks cluster management systems has been adapted in GEON for wide-area operation. GEON provides the standardization and convenience that comes with a centralized software stack definition, but has distributed "Centrals" (a "Central" is a site that has oversight for an individual component of the software stack).

Data Management. GEON supports the notion of "hosted" and "nonhosted" datasets. Hosted datasets are physically stored at a centralized location, along with their metadata records. Non-hosted datasets are stored at distributed locations—the corresponding metadata records may be stored centrally, or even distributed. A data registration capability at the GEON portal supports the ability to search and discover data in GEON using a variety of spatial, temporal, and attribute-based search conditions. GEON is currently experimenting with the Globus Data Replication Service (DRS) to manage data replication in the GEON distributed environment.

Interoperability. While the system incorporates a main portal, there is also the need to support distributed portals as well as remote portlets. The distributed portals share common systems services (for example, authentication). Remote portlets may be invoked from any of the distributed portals. The system is based on a service-oriented architecture but needs to support interoperability between .NET and AXIS-based services.

Support for End-User Environments. The GEON Web portal is based on the GridSphere portal infrastructure. Features have been added to GridSphere to integrate Flash applications, as well as to support distributed portals that allow individual sites to customize their portal functionality and to "brand" their portals. Support for user environments also includes providing a scientific workflow capability, which is done in GEON via the Kepler workflow tool.

Data Integration and Mediation. Advanced data integration is facilitated by the use of ontologies to integrate data from heterogeneous schemas and databases. The GEON data integration middleware, or mediator, employs a Simple Ontology-Based Query Language (SOQL) to facilitate such data integration. In the process, tools have also been developed to register data items and schemas to ontologies.

Grid Account Management Architecture 2.0

By Karan Bhatia[1], Sandeep Chandra[1], and Kurt Mueller[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

Security is a critical component of grid systems and numerous software components and tools provide some security capabilities; however, few end-to-end security systems for emerging grid infrastructures work "out-of-the-box." Hence, significant time and effort are spent evaluating software and testing and integrating different software components, even though most grid infrastructures have very similar requirements and needs. Grid Account Management Architecture (GAMA) 1.0 addressed this need and provided a solution packaged as part of the Rocks cluster management software. This software is in production use by the Geosciences Network (GEON), the Biomedical Informatics Research Network (BIRN), the Telescience Project at University of California—San Diego, Indiana University, Harvard University's Crimson Grid, and the Pacific Rim Applications and Grid Middleware Assembly (PRAGMA), among others. This poster describes the next major revision of GAMA, which features (1) better support for distributed infrastructures, (2) a more refined administrative console with the capability to add custom plug-ins; (3) support for both certificate authority client (CACL) and National Research Grid Initiative Client Authority (Naregi CA) software; (4) integration with cluster accounts; and (5) a basic authorization mechanism.

GAMA 2.0 consists of two components: a server component (the task engine) that creates and manages X.509 user credentials on behalf of the user and a set of portal components that provide interfaces for both users and administrators. The core functionality of the server is made available by a set of application programming interfaces (APIs). These APIs allow clients to request authentication and authorization, request that new users be created and added to the system, and retrieve the public information of individual users. These functions are grouped into three types: a user API, a public administration API, and a private administration API. The user API is used by clients on behalf of users to authenticate an individual or provide authorization information for a user. The public administration interface provides a means for such clients as portals or cluster nodes to retrieve public information about users and to submit user account requests. The private administration interface provides the means to accept users, create new resources, and update the authorization of any user. This last interface must be private and secured; otherwise, the entire system could become compromised. The public interfaces are available as Web-service APIs with associated port types and Web Services Description language (WSDL) documents to describe them. The private interface may eventually be implemented as a secure Web-service port type; however, in this design, private administration only is provided through a Java library and, hence, is not available from remote sites. As mentioned previously, the administration functions are provided through a Web interface using a server running on the GAMA server itself. The public interfaces are run in a secure container and require Hypertext Transfer Protocol Secure (HTTPS)-encrypted communication.

The design of the server component is "implementation agnostic," meaning that there is no explicit reliance on any back-end components. This is accomplished through a plug-in architecture in which administrators choose a CA, credential storage, authorization service, and so forth, of their choice. The architecture makes server operations customizable and expandable through a workflow-type system that can execute arbitrary sequences of tasks for any operation. The architecture also supports multiple sites on one server; each site has local administrators responsible for managing user accounts for the site. Users may be given access from one or multiple sites. This feature can be configured using the portal components.

GAMA 2.0 provides account management for clusters by allowing administrators to determine which users have access to the clusters on the centrally managed GAMA server. The cluster must have GAMA 2.0 client support and accounts must be synchronized periodically based on the information provided by the server. The head node for the cluster must be installed with the GAMA 2.0 client (Rocks roll) software and configured at the time of cluster installation. This configuration process consists of (1) specifying a GAMA server host name that will manage the accounts, and (2) specifying a "resource_id" and "resource_pwd", which are defined at the GAMA server prior to cluster installation. A new group "gama" is created on the head node under which all accounts will be created, and the host name, "resource_id" and "resource_pwd" are stored on the cluster front end. The client software is configured to synchronize with the server every 24 hours as a cron (automatically scheduled) job.

GIS Activities in the USGS Energy Resources Program—A Model for GIS Utilization

By Laura R.H. Biewick[1], Gregory L. Gunther[1], Christopher C. Skinner[1], and David A. Ferderer[1]

[1]U.S. Geological Survey, Denver, Colo.

A primary objective for the implementation of Geographic Information System (GIS) technologies by the Energy Resources Program (ERP) of the U.S. Geological Survey (USGS) is to improve access to maps, data, and other geospatial services. The effort, in turn, is intended to improve the capability of decisionmakers to analyze layers of disparate data in a common geographic space.

Use of GIS technologies by the ERP is enhancing research activities related to project workflow and information access and discovery by providing the following: (1) efficient centralized data management and data visualization, (2) ease in sharing data and interpretations among project personnel, and (3) dissemination of information and products to customers in an easily usable format.

ERP GIS activities include Internet map services and metadata services, which are also being leveraged in global networks that provide the infrastructure needed to support the sharing of geographic information. These portals include the National Spatial Data Infrastructure, the Geosciences Network (GEON), and the GeoSpatial One-Stop. Major tasks include technical issues related to application deployment, security, and system architecture. Demonstrations of the National Oil and Gas Assessments (NOGA) at NOGA Online, Gulf Coast Geology (GCG) at GCG Online, the Gulf Coast Information Access System, and World Energy Assessment applications illustrate how interactive maps and publication services provide easy access to organized assessment results, supporting geologic studies and other ERP project data and interpretations.

SCEC/CME CyberShake—Probabilistic Seismic Hazard Analysis Using 3D Seismic Waveform Modeling

By Scott Callaghan[1], Philip Maechling[1], Yifeng Cui[2], Marcio Faerman[2], Ned Field[3], Robert Graves[4], Nitin Gupta[1], Vipin Gupta[1], Thomas Jordan[1], Carl Kesselman[5], Gaurang Mehta[5], David Okaya[1], Karan Vahi[5], and Li Zhao[1]

[1]University of Southern California, Los Angeles, Calif.

[2]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[3]U.S. Geological Survey, Pasadena, Calif.

[4]URS Corporation, Pasadena, Calif.

[5]University of Southern California Information Sciences Institute, Marina Del Rey, Calif.

Researchers from the Southern California Earthquake Center Community Modeling Environment (SCEC/CME) project are utilizing the CyberShake computational platform to calculate probabilistic seismic hazard curves for many sites in the Southern California area. Traditionally, probabilistic seismic hazard analysis (PSHA) is conducted using intensity measure relationships (IMRs) based on empirical attenuation relationships. These attenuation relationships represent relatively simple analytical models based on the regression of observed data. However, a more physics-based approach using waveform modeling could lead to significant improvements in seismic hazard analysis. Therefore, the CyberShake hazard curves use IMRs based on three dimensional (3D) ground motion simulation, rather than on attenuation relationships.

To introduce 3D seismic waveform modeling into PSHA hazard curve calculations, the SCEC/CME CyberShake group has integrated (1) a leading-edge PSHA software tool (OpenSHA), (2) an SCEC-developed geophysical model (SCEC CVM3.0), (3) validated anelastic wave modeling (AWM) software, and (4) state-of-the-art computational technologies, including high-performance computing and grid-based scientific workflows, to develop an OpenSHA-compatible, 3D, waveform-based IMR platform. The integration of this technology will allow researchers to combine a new class of waveform-based IMRs with the large number of existing PSHA components, such as earthquake rupture forecasts (ERFs), which are currently implemented in the OpenSHA system.

To calculate a probabilistic hazard curve for a site of interest, we use the OpenSHA implementation of the National Seismic Hazard Mapping Project (NSHMP)-2002 ERF and identify all ruptures within 200 kilometers of the site of interest. For each of these ruptures, we convert the NSHMP-2002 rupture definition into one or more ruptures with slip time history (rupture variations) using newly developed rupture-generation software. Strain Green tensors are calculated for the site using well-validated AWM software together with the SCEC CVM3.0 3D velocity model. Then, using a reciprocity-based approach, we calculate synthetic seismograms for each rupture variation. The resulting suite of synthetics is processed to extract peak intensity measures of interest (such as spectral acceleration). The peak intensity measures are combined with the original rupture probabilities to produce probabilistic seismic hazard curves for the site.

The CyberShake calculations require thousands of central processing unit (CPU) hours and several terabytes of disk storage. They are performed, therefore, on high-performance computing systems, including multiple TeraGrid sites (currently San Diego Supercomputer Center (SDSC) and National Center for Supercomputer Applications (NCSA)) and the University of Southern California's High Performance Computing and Communications (HPCC) center. CyberShake utilizes a grid-based scientific workflow system based on the Virtual Data System (VDS) to manage the extensive job scheduling and data management requirements of the project.

Hazard curves for the initial 10 sites, including rock and basin sites, have been completed. Computation for additional sites is proceeding with the aim of creating a physics-based seismic hazard map for the Southern California region.

Toward Formal Web Service Behavioral Descriptions

By Nicholas Del Rio[1]

[1]The University of Texas—El Paso, El Paso, Tex.

Because of the black box nature of Web services, their development and consumption pose issues that are not present in classical distributed environments: the developer does not know the context in which the service will be used, and the consumer does not know how a service component was developed. This ambiguity can result in consumers using Web services incorrectly. For example, a geoscientist might run an application that consumes a Bouguer anomaly Web service, but this service may operate with a lower precision than the geosciences application requires. Formal properties specified in a Web service descriptor could have unambiguously indicated that this Web service works with a lower precision.

Somehow, these run-time properties must be communicated to the consumer using some kind of descriptor. Currently, the Web Services Description Language (WSDL) documents serve as the fundamental resource for describing Web services; however, WSDL supports only the basic constructs necessary for binding a Web service, such as supported operations, request/response message formats, network protocol and endpoints. WSDL lacks support for formally expressing Web service behavioral properties. This research proposes solutions to enrich WSDL or some other service descriptor with the ability to express formal software properties.

Internet Technologies in GIS-Based Earth Science Applications

By Asli Garagon Dogru[1], Gonul Toz[2], and Haluk Ozener[1]

[1]Kandilli Observatory and Earthquake Research Institute, Geodesy Department, Bogazici University, Istanbul, Turkey

[2]Civil Engineering Faculty, Photogrammetry Division, Istanbul Technical University, Istanbul, Turkey

The Internet has moved beyond its original intent. Today, it is possible to design conventional geographic information system s (GISs) that can share various data over the Internet use of the whole world. Internet GIS applications also have become education tools for the public to learn about the importance of GIS for understanding of the Earth.

In Turkey, there are many problems related to Internet GIS in the earth sciences. The most important one is that geoscientific data from earthquake studies are stored and processed using various software applications. Furthermore, large amounts of these data are available in different forms and projections. They come from various resources and they do not have a datum defined; however, an integrated system can be built by combining these geoscientific datasets into a common GIS platform, describing the standardization methods used, and sharing them over the Internet. In this way, a learning environment of GIS-based earthquake applications over the Internet would be provided for the public as well.

This study aims to investigate how the Internet can bring the benefits to data integration and data distribution to users who need efficient ways of passing spatial data to and from different locations. Another purpose of this study is to describe and analyze the status of data sharing and distribution in institutions that produce geospatial data on a national scale in Turkey.

Beowulfs' Place in the Computational Infrastructure for Geographic Modeling

By David I. Donato[1]

[1]U.S. Geological Survey, Reston, Va.

Grids, virtual machines, and other classes of high-performance computing aggregates may well prove to be more suitable than Beowulf clusters for many tasks in geographic modeling (and in scientific computation in general). Even so, Beowulfs still provide a better balance of throughput, lower unit computational cost, and better security for many of the tasks encountered in geographic modeling than these other computing resources if idle time is kept low, job scheduling is efficient, and administrative costs are contained through widespread sharing of experience and software among organizations that operate Beowulfs. We briefly survey the current state of the economics of scientific computation, propose the application domains in geographic modeling for which Beowulfs will continue to be suitable, and relate our experience in developing an Internet interface in order to provide secure, distributed access to a Beowulf dedicated to geographic computation and modeling.

The Use of GIS to Support Atmospheric and Oceanographic Data Management and Mapping

By Fan Hong[1,2] Di Liping[1], Zhao Qifeng[2], and Lai Jianfei[2]

[1]Laboratory of Advanced Information Technology and Standards, George Mason University, Fairfax, Va.

[2]National Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

As the ocean becomes increasingly important to human activities, sustainable exploitation and use of ocean resources have become the focus of many research projects. With the progress on ocean observing systems and the increase of ocean monitoring activities, a large amount of ocean data have been collected and many analysis results have been obtained. All these challenge the traditional file-based data management and manual or semiautomatic mapping methods. The traditional methods can barely accomplish comprehensive management, analysis, and applications of dynamic ocean data because of the spatial and temporal complexity of atmospheric and oceanographic data and information. Geographic information systems (GIS) can provide powerful functions to support comprehensive management of spatial data; these functions include storage, processing, analysis, and visualization. Such functions have been successfully applied in many fields, such as traffic management and urban planning. Introducing GIS technology and methods to the ocean fields can help achieve the goal of comprehensive management of dynamic and complex atmospheric and oceanographic data and information. However, because most GISs have been developed for land applications, they cannot be directly used in ocean applications.

This paper introduces the design and development of a GIS for ocean applications—the Xiamen Atmospheric and Oceanographic Data Management and Display System (AODMDS). The system is based on ArcObjects (AO), which is a component-based GIS development tool. The paper discusses in detail the storage and organization of the atmospheric and oceanographic data, the strategy and methods for visualization and mapping of oceanographic and atmospheric data, and the implementation of the methods in AODMDS. It also discusses advanced display control techniques that expand the functions of AO. An example of such techniques is the "gradient-fill-style color-map control," which provides a feasible color-rich display control for all types of raster maps. AODMDS was successfully deployed in December 2005 at the Fourth Institute of Oceanography, China, to replace the old data management and mapping system. The operation of the system shows that AODMDS can support effective mapping and visualization of oceanographic data and information. As a standalone desktop GIS system built on AO, AODMDS has achieved the effective data management and powerful mapping and visualization of atmospheric and oceanography data. The performance of the system has met the requirements and goals of the ocean user community very well.

An Approach for Identifying and Specifying Runtime Monitoring Properties for Geoinformatics Software Applications

By Irbis Gallegos[1] and Nicholas Del Rio[1]

[1]University of Texas—El Paso, El Paso, Tex.

Some geosciences software applications can require intensive computational loads. For example, a seismic tomography algorithm may run for many hours, depending on the initial velocity model, parameters, and number of iterations requested by the user. Much time may pass before the user is aware that the application is not behaving as expected. To show the scientist that the velocity models are converging, it would be desirable to add a checkpoint that signifies "step completed to update slowness factor for each cell and the model has been smoothed;" and to add a property that checks the following at that checkpoint: "the RMS of the travel time residuals are less than or equal to the pick error." Software monitoring can support this scenario by checking the application as it runs against predefined properties. For this to work, however, it is necessary to identify observable entities in the software application, define the properties to be monitored on those entities, and provide a mechanism for communicating the properties to the consumer.

To address the first problem, we propose a new approach that identifies observable items based on monitoring patterns derived from use case models. Most of the existing monitoring systems assume that developers are software experts with domain expertise; however, in practice, this is not the case.

To address the second problem, we propose to enrich the Web Services Description Language (WSDL) or some other service descriptor with the ability to formally express the identified items and properties. Currently, WSDL documents serve as the fundamental resource in which Web services are described; however, WSDL supports only basic constructs necessary for binding a Web service and lacks support for formally expressing Web service behavioral properties. This poster presents our initial results in addressing this problem.

Development of a GIS-Based Regional Upper Mississippi Valley Karst Feature Database

By Yongli Gao[1], Emmit C. Alexander, Jr.[2], and Michael J. Bounk[3]

[1]Department of Physics, Astronomy, and Geology, East Tennessee State University, Johnson City, Tenn.

[2]Department of Geology and Geophysics, University of Minnesota, Minneapolis, Minn.

[3]Iowa Geological Survey, Iowa City, Iowa

This paper presents the ongoing effort to develop a multistate Upper Mississippi Valley Karst Feature Database (KFD) using geographic information system (GIS) and database management system (DBMS) technologies. A regional KFD of the western part of the Upper Mississippi Valley Karst, which includes southeastern Minnesota and northeastern Iowa, was created by merging the Minnesota KFD with available karst-feature and geologic information in Iowa. Karst-feature, geologic, and geographic datasets in Wisconsin and Illinois are currently being assembled into the database.

Existing State, county, and sub-county karst-feature datasets in Minnesota and Iowa have been assembled into the KFD, which is capable of visualizing and analyzing the entire dataset across State boundaries. GIS coverages of all the datasets were projected based on the 1983 North American Datum (NAD). The western side of the Upper Mississippi Valley Karst is subdivided into the Prairie du Chien Karst, the Galena-Maquoketa Karst, and the Silurian-Devonian Karst. This KFD will provide data and tools for researchers and land-use planners to access karst hazards and to manage water resources of the Upper Mississippi Valley Karst. The ultimate goal of this research is to develop a complete regional Upper Mississippi Valley KFD that includes all karst features and relevant geologic and geographic information in northwestern Illinois, southwestern Wisconsin, northeastern Iowa, and southeastern Minnesota.

A Three-Tier Architecture for LiDAR Interpolation and Analysis

By Efrat Jaeger-Frank[1], Christopher J. Crosby[2], Ilkay Altintas[1], Viswanath Nandigam[1], J. Ramon Arrowsmith[2], Jeffery Conner[2], Ashraf Memon[1], and Chaitan Baru[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[2]Department of Geological Sciences, Arizona State University, Tempe, Ariz.

The Geosciences Network (GEON) is a large-scale Information Technology Research (ITR) project, funded by the National Science Foundation, which facilitates a collaborative, interdisciplinary effort in the field of earth sciences.

Components of the GEON cyberinfrastructure include advanced semantic-based discovery and integration of data and tools via the GEON portal. The GEON portal provides unified and authenticated access to a wide range of resources, which allow geoscientists to conduct comprehensive analyses using emerging technologies in order to facilitate the next generation of science and education. One of the many challenges in GEON is the efficient management and analysis of the LiDAR (Light Distance And Ranging) point cloud datasets. For most geosciences users, the sheer volume and (high) density of these point clouds make distribution and interpolation of the LiDAR data difficult. Furthermore, efficient manipulation and analysis of these massive volumes of data cannot be achieved with a conventional data distribution and processing system.

In this poster, we present a scientific workflow approach to LiDAR data interpolation and analysis, coordinating various resources as data analysis pipelines, through a three-tier architecture. The architecture utilizes the GEON portal as a front end user interface, the Kepler scientific workflow engine, and the Grid infrastructure. Although the example is focused on the LiDAR data processing, it can be used as a model in other problem areas with similar requirements.

Search, Access, and Visualization of 3D Volume Data on the GEONgrid Portal

By Ashraf Memon[1], Ghulam Memon[1], and Chuck Meertens[2]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[2]University NAVSTAR (Navigation Signal Timing and Ranging) Consortium (UNAVCO), University of Colorado, Boulder, Colo.

The Geosciences Network (GEON), which is funded by the National Science Foundation (NSF), undertakes the task of integrating and visualizing geosciences data in multidimensional space.

Integration and visualization of three-dimensional (3D) data involves several challenges, including those related to the size of the data and the format and tools for visualization. GEON is aiming to provide the portal-based framework for search, access, and visualization of 3D volumes from distributed sources. Network common data form (netCDF) is one of the most popular formats for storing and transmitting volume data. Hence, our first milestone in this rough terrain will be the complete and seamless visual integration of netCDF, either hosted or nonhosted, with GEON's existing framework.

In this poster, we address the use of service-oriented architecture for registration, search, and on-demand visualization of the 3D volume data from Earth and atmospheric sources (for example, seismic tomography and atmospheric temperature) on the GEONgrid portal. A variety of protocols and standards are addressed; they include the Open-Source Project for a Network Data Access Protocol (OPeNDAP), Thematic Realtime Environmental Data Distributed Services (THREDDS), Open Geospatial Consortium (Web Map Service/Web Feature Service/Web Coverage Service (WMS/WFS/WCS) and Geo-Interface to Atmosphere, Land, Earth, Ocean NetCDF (GALEON)), netCDF, International Organization for Standardization (ISO), and Open Archive Initiative. A proposed solution is discussed. We also consider the registration of OPeNDAP/THREDDS service into GEON and ways to exchange the appropriate metadata about the service.

Extending Akamai to Incorporate Application-Specific P2P Overlays

By Ghulam Memon[1] and Chaitan Baru[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

Akamai Technologies, Inc. introduced the phenomenon of Web caching by deploying an overlay network of approximately 10,000 caches around the world. The basic idea was to cache static images and other embedded Hypertext markup Language (HTML) content at strategic points in the network, so that the domain name server (DNS) could facilitate efficient retrieval of the content.

Close examination of Akamai's architecture reveals that it has a centralized architecture; that is, the logical location of nodes can be obtained from the DNS and brought down by a malicious user. Moreover, the Website-centric nature of Akamai's model makes it less flexible and attractive for interactive Web applications, such as geographic information systems (GISs). Also, building such fine-tuned predetermined overlays, which are specific to applications, is not an option because of costs.

In this poster, we present a dynamic peer-to-peer (P2P)-based Web-caching architecture that is specific to but not limited to GIS applications, that can remove the above-mentioned problems without losing the benefit of efficiency. Recent proposals for true P2P infrastructures, such as Pastry, Tapestry, Chord and content-addressable networks (CAN), promote fully decentralized architectures. Hence, using one of these infrastructures solves the problem of centralization. We chose Pastry for our implementation because of its inherent property of locality. Pastry also facilitates application-specific overlays without incurring the above-mentioned costs.

The technique demonstrated by this poster involves hosting data and pointers to remote data services on GEON partner sites, which will act as both bootstrap nodes and super nodes (because they host original data or act as GEON's gateway to distributed data sources) and will hash the search criteria on existing peers. The data can be discovered by hashing the queries and then routing to the closest possible cache or super node. The poster will also demonstrate how to use a BitTorrent-style algorithm to divide data into n-dimensional pieces so that the multidimensional nature of spatial data can be exploited.

Seamless Integration of Heterogeneous Data Sources

By Ghulam Memon[1] and Ashraf Memon[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

The Geosciences Network (GEON) is a cyberinfrastructure project funded by the National Science Foundation (NSF). GEON is developing infrastructure for knowledge-based integration of geosciences data. At its heart, GEON is aimed at collecting heterogeneous spatial data (based on its data format) from different sources, either hosted or nonhosted, and then developing tools and techniques to analyze and integrate these data.

Heterogeneous data sources pose a technical challenge of integration, despite their potential semantic relationship. In this poster, we present a generic scheme by which scientists will have the ability to overlay different spatial images on top of each other without worrying about the physical file format. This work goes beyond simple layering of Graphics Interchange Format (GIF) and Joint Photographic Experts Group (JPEG) images. From a very high level, the proposed component will convert different data formats into one uniform bundle on the fly and then use existing GEON tools for integration and visualization.

For example, shapefiles, georeferenced raster imagery (GeoTIFFs), and all formats supported by Environmental Systems Research Institute (ESRI) can be integrated by using ArcIMS, and ESRI image service. Databases and spreadsheets, on the other hand, can be integrated together and with ESRI file formats by creating temporary shapefiles based on their spatial component. All ESRI file formats can be integrated with remote Web Map Service (WMS) layers by converting them into WMS data sources that are hosted locally. All ESRI-supported formats can also be integrated with Network common data form (netCDF) files by converting them into WMS services and using the Geoscience Network Interactive Data Viewer (GEON IDV) for overlaying with three-dimensional volumes stored in NetCDF files.

Upon implementation of this technique, GEON users will be able to add spatial resources in the integration cart and create a combined visual representation of the selected datasets.

Building a Cyberinfrastructure for the Critical Zone Exploration Network

By D.A. Miller[1], S.L. Brantley[1], M.S. Hofmockel[2], and D.D. Richter[2]

[1]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pa.

[2]Duke University, Durham, N.C.

This investigation will initiate a cyberinfrastructure to promote interdisciplinary research on the zone defined by the outer limits of vegetation and the lower boundary of ground water—an area identified by the National Research Council as "the Critical Zone." A complex interplay of physical, chemical, and biological reactions controls processes within the Critical Zone and has become the focus of research among a community of scientists from the fields of ecology, soil science, biology, geochemistry, hydrology, and geomorphology. The Critical Zone Exploration Network (CZEN) is envisioned as a network of sites, people, tools, and ideas that will (1) investigate the response of the Critical Zone to global change, (2) promote systematic data and sample collection, (3) provide databases and sample storage facilities, and (4) train the next generation of Critical Zone scientists.

The CZEN cyberinfrastructure must meet the needs of scientists from an array of disciplines and institutions. Many of these disciplines have their own conceptualizations, nomenclature, and methodologies for weathering processes. This research project will develop an ontology, or integrative language, for Critical Zone science that will serve to bridge discipline boundaries. Building this ontology through a community-based effort will not only produce a focused, highly integrative communication standard for advancing the science of the Critical Zone, it will foster new collaboration and partnerships among groups that have not traditionally interacted. The development of this cyberinfrastructure will centralize Critical Zone data and metadata and increase the usefulness of Critical Zone studies. A Web-based knowledge-management environment to facilitate collaboration and sharing of data collected worldwide will be created to meet these goals.

This project will contribute directly to the developing CZEN initiative by fostering interaction among the wide range of scientists and scientific disciplines that study the earth's Critical Zone. The project will also contribute to the broader goals of developing a cyberinfrastructure for the geosciences through the existing Geoscience Network (GEON) program. By linking with efforts in GEON and other ongoing initiatives within Federal agencies and academic groups, the CZEN cyberinfrastructure will build upon knowledge from personnel throughout the geosciences in efforts to develop the required electronic infrastructure. Ultimately, this effort will itself contribute to the vision of a comprehensive geoscience cyberinfrastructure that underpins the GEON program.

Opening Doors for Seismic Data Access

By Joanna Muench[1], Linus Kamb[1], Rob Casey[1], and Tim Ahern[1]

[1]Incorporated Research Institutions for Seismology (IRIS) Data Management Center, Seattle, Wash.

The distribution of seismic data remains a key responsibility of the Incorporated Research Institutions for Seismology (IRIS). Over the last 20 years, the means of providing access to data has shifted with technological changes, from sending tapes via surface mail to high-speed data transfers. An increasing priority is enabling programmatic access to data and analysis tools, and providing capability for members of the seismic community to share a wider variety of data products. Adoption of a service-oriented architecture at IRIS is a critical means to achieving these goals.

IRIS initiated our service-oriented approach five years ago with the introduction of the Common Object Request Broker Architecture (CORBA)-based data-handling interface (DHI). The DHI provides programmatic access to multiple seismic data centers through Java and C/C++ clients. More recently, the commonly used Seismic Analysis Code (SAC) program was updated to interface with DHI, bringing data access and analysis tools together. Noting the maturation of Web services into an accepted data service technology, IRIS has developed a prototype Web service version of the DHI using the Simple Object Access Protocol (SOAP).

In the near future, IRIS will release DHI 2.0. This new version will encompass both CORBA and SOAP technologies to provide users with data access technologies most appropriate for their needs. Also in development is the Searchable Product Archive and Discovery Engine (SPADE), which is a tool to allow registered data producers from locations across the globe to make their products available to the community. With SPADE, scientists will be able to discover data sources through a single SOAP interface, searching via spatial, temporal, and product-specific metadata.

As we move forward, IRIS will continue to integrate and provide access to its data and analysis tools programmatically, as components of Web sites and as standalone applications. Web service and portlet technologies allow easy incorporation of IRIS tools and data into research programs, educational Web sites and Government planning. We are also exploring data mining, data presentation, and mapping tools as further contributions to the growing field of cyberinfrastructure development.

Construction of an e-Science Environment for the Weather Information System

By Jai-Ho Oh[1]

[1]Department of Environmental Atmospheric Sciences, Pukyong National University, Busan, Korea

Atmospheric data production systems are now turning out petabytes of data for earth systems research. To manage this massive dataset, advanced tools are necessary. An improved research environment for progressive meteorological research is required, as is the development of a collaborative system for meteorological research and computational resources. In the research described in this poster, the main objective of this research is to construct an atmospheric e-science environment for research activities.

In order to accomplish this objective, we took a four-step approach. First, we employed the Globus Toolkit components. Second, we devised the methods of interoperation between the meteorology databases. Third, we defined the workflow to create meteorological metadata. Finally, we devised the meteorology prediction system configuration and the display system based on Web interfaces.

The Meteo-DataGrid system is designed for integration of distributed data and efficient metadata cataloging. Raw meteorological data is obtained from different types of data servers and is accumulated in a sequential-access-to-metadata (SAM) file pool. The Center for Ocean-Land-Atmosphere Studies' (COLA's) Grid Analysis and Display System (GrADS) Data server (GDS) and the Live Access Server (LAS) hosted by the National Oceanic and Atmospheric Administration are the most widely used servers in international meteorological research and they were used as our data servers. The metacatalog for the various forms of weather datasets was prepared using Data Broker Interface, and the data pool was constructed using this metacatalog. The data pool was next coupled with GridSphere with GrADS graphical display system of weather data. Through this atmospheric e-science environment, meteorological researchers can obtain the necessary data and can visualize these datasets using GridSphere as their Web-based interface. The research community can have easy access to datasets and they can acquire technical knowledge irrespective of differences in datasets.

As a part of the e-science environment, a regional-scale weather model, MM5, that uses a Web interface is currently in operation. Plans are in place to operate various global climate system models using a Web-based interface. Plans are also in place to revitalize the research group using computational grid techniques (by Unicore and Globus) for the use of earth system modeling. This improved e-science environment will play an important role in the collaborative research of atmospheric sciences.

Talwani—A Profile Modeling Tool Adaptable to Cyberinfrastructure

By Julio Olaya[1], Leonardo Salayandía[1], Ann Gates[1], and G. Randy Keller[2]

[1]Department of Computer Science, University of Texas—El Paso, El Paso, Tex.

[2]Department of Geological Sciences, University of Texas—El Paso, El Paso, Tex.

Among the goals of cyberinfrastructure efforts is ready access to data and the software to process and model them. In the case of measurements of the Earth's gravity and magnetic fields, recent efforts have produced major new databases for North America. However, modern open-source modeling software is needed. In response to this need, we have developed the Talwani software tool, which is a profile forward modeling application based on the technique of Talwani and others (1959) and Cady (1980). The previous version of the program (version 2.2) allowed the creation of forward models of gravity profiles; the present version, version 3.0, adds the functionality to create magnetic models. Another enhancement to the Talwani software tool is support to input models in Extensible Markup Language (XML). Supporting a portable XML format will allow the models to be annotated with semantic and provenance information, which will facilitate knowledge sharing, reuse, and determination of trust. As developers of the tool, we are applying the "design for change" software engineering principle in anticipation of changes as to how the models produced by the tool will be used. We envision the use of cyberinfrastructure efforts such as the Geosciences Network (GEON) to share and disseminate models among the scientific community, and to facilitate model integration to other resources.

References Cited

Cady, J.W., 1980, Calculation of gravity and magnetic anomalies of finite-length right polygonal prisms: Geophysics, v. 45, p. 1507-1512.

Talwani, M., Worzel, J.L., and Landisman, M., 1959, Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone: Journal of Geophysical Research, v. 64, p. 49-61.

The SCEC TeraShake Simulations—High-Resolution, Regional-Scale Simulations of Large Southern San Andreas Earthquakes

By Kim Olsen[1], Steven Day[1], J. Bernard Minster[2], Yifeng Cui[2], Amit Chourasia[2], Reagan Moore[2], Yuanfang Hu[2], Jing Zhu[2], Philip Maechling[3], Thomas Jordan[3], and David Okaya[3]

[1]Department of Geological Sciences, San Diego State University, San Diego, Calif.

[2]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

[3]Department of Earth Sciences, University of Southern California, Los Angeles, Calif.

Researchers working on the Southern California Earthquake Center Community Modeling Environment (SCEC/CME) Project have carried out some of the largest and most detailed earthquake simulations completed to date (TeraShake). The researchers modeled ground motions expected from a large earthquake on the southern San Andreas fault on parallel supercomputers using facilities at the San Diego Supercomputer Center (SDSC) and National Center for Supercomputer Applications (NCSA).

The TeraShake calculations simulate 4 minutes of 0 to 0.5 Hz (hertz, or cycles per second) ground motion in a 180,000 square kilometer area of southern California, for a magnitude (M) 7.7 earthquake along the 199-kilometer-long section of the San Andreas fault located between Cajon Creek north of Los Angeles and Bombay Beach on the shore of the Salton Sea. The two segments of the San Andreas fault south of the 1857 rupture (the San Bernardino Mountains segment and the Coachella Valley segment) have not seen a major event since 1812 and about 1690, respectively. The average recurrence interval for large earthquakes with surface rupture on these segments is only 146+91-60 yrs and 220±13 years, respectively. An inescapable conclusion is that a major component of the seismic hazard in southern California and northern Mexico stems from a putative large earthquake on this part of the San Andreas fault.

The TeraShake simulations include ruptures that propagate both northwestward and southeastward on the fault. The TeraShake simulations use both kinematic (based on that inferred for the 2002 Denali, Alaska, earthquake) and dynamic (based on the 1992 Landers, Calif., earthquake) source models. The crustal model is taken from the SCEC Three Dimensional Community Velocity Model Version 3.0 discretized into 200-cubic-meter (cm^3) cubes.

The results show that the chain of sedimentary basins between San Bernardino and downtown Los Angeles form an effective waveguide that channels Love waves along the southern edge of the San Bernardino and San Gabriel Mountains. Earthquake scenarios in which the guided wave is efficiently excited (scenarios with northward rupture) produce unusually high long-period ground motions over much of the greater Los Angeles region. Intense, localized amplitude modulations arising from variations in waveguide cross-section can be explained to a remarkable level of accuracy in terms of energy conservation for the guided mode.

While the kinematic and dynamic source models generate similar patterns of peak ground motions, the peak ground motions for the dynamic model are generally somewhat smaller than those for the kinematic source. The main reason for the latter finding is the primarily less-coherent wavefronts generated by the more-complex dynamic rupture history. Adding to the uncertainty of the predicted absolute amplitudes of the ground motion extremes is nonlinearity induced by the higher-than-anticipated waveguide amplifications we have identified, which would likely cause significant reduction of both shear modulus and Q factor in the near-surface layers. Animations of the simulated wave propagation and synthetic seismograms from TeraShake can be found online at http://www.scec.org/TeraShake.

Believing Answers From GEON Applications

By Gilbert Ornelas[1] and Paulo Pinheiro da Silva[1]

[1]Department of Computer Science, University of Texas—El Paso, El Paso, Tex.

The Geosciences Network (GEON) aims to increase collaboration between applications (such as services and tools) and to increase the reuse of information sources by applications. This suggests that application results might be coming from multiple sources and services may unintentionally provide information that contains discrepancies or is inaccurate. Thus, it is becoming important for GEON applications to have some kind of tracking mechanism to indicate from where results were derived (that is, provenance information). In addition to provenance information, the applications may need a mechanism for computing belief recommendations for their results that are based on a web of trust that GEON users may have developed for sources; such a web of trust might be able to identify the users' degrees of trust of sources as well as the user's trust of other users. Trust recommendations may be used by users to build opinions about the trustworthiness of the results themselves and of any intermediary result provided during the processes of generating answers. In this poster, we present initial results of our effort to develop and integrate a trust component and infrastructure within GEON's service-oriented architecture. By using provenance information and trust recommendations, we expect geoscientists to easily decide how much to believe results produced by applications using a complex set of GEON services and information sources. As a result, geoscientists will have more control over which results they may accept with confidence and which results should be further questioned.

Ontology Re-engineering Use Case—Extending SWEET to Map Climate and Forecasting Vocabulary Terms

By Rahul Ramachandran[1], Sara Graves[1], and Rob Raskin[2]

[1]Information Technology and Systems Center, University of Alabama—Huntsville, Ala.

[2]Jet Propulsion Laboratory, Pasadena, Calif.

A common problem faced while developing metadata for scientific data archives is that of keywords. Although keywords are an effective way for searching the resource catalogs, data archive designers may select from one of many different controlled vocabularies to describe their holdings. For example, in the earth sciences, the Climate and Forecasting (CF) Convention is a controlled vocabulary commonly used within the modeling community. Similarly, the Global Change Master Directory (GCMD) keywords list is the convention used within the Earth Science Program at the National Aeronautics and Space Administration. The use of controlled vocabularies allows searches on the resource catalogs to be accurate and complete, but the burden of framing the precise query falls on the shoulders of the user. The user has to know the keyword beforehand to perform a search, which might be perfectly acceptable in smaller projects in which the user is specialized and has the required knowledge, but is impractical in larger projects in which the user may have varied levels of domain knowledge. One solution to this problem is the use of an ontology that contains higher-level abstract concepts that are mapped to the different controlled vocabulary terms. This use of ontologies eliminates the barrier of entry based on domain knowledge and provides easy-to-use search capabilities to the users. In this poster, we describe an ontology designed and created to address this problem. However, this ontology required re-engineering of higher-level ontologies, namely the Semantic Web for Earth and Environmental Terminology (SWEET) ontologies, instead of the creation of an original ontology. Because the traditional methods of creating an ontology do not account for reengineering and reuse of higher level ontologies, we propose a new modified methodology. This presentation describes this methodology and explores some of the issues and challenges involved in the construction of an ontology using this approach.

Formalization of Fold and Fault Concepts in a Structural Geology Ontology

By Stephen M. Richard[1], Hassan A. Babaie[2], John S. Oldow[3], Hans G. Avé-Lallemant[4], and A. John Watkinson[5]

[1]Arizona Geological Survey and U.S. Geological Survey, Tucson, Ariz.

[2]Department of Geosciences, Georgia State University, Atlanta, Ga.

[3]Department of Geological Sciences, University of Idaho, Moscow, Idaho

[4]Department of Earth Science, Rice University, Houston, Tex.

[5]Department of Geology, Washington State University, Pullman, Wash.

Development of structural geology ontologies for use by geologists and users of the Geosciences Network (GEON; www.geongrid.org) was initiated in October 2004 when a group of domain experts met at the University of Idaho in Moscow, Idaho. Challenges facing the development of a computable knowledge representation for structural geology include defining the scope and purpose of the structural geology ontologies and formalizing definitions of the common domain terminology used to describe geologic structures. Although there is no disagreement that faults and folds should be included in a structural geology ontology, knowledge boundaries are fuzzy at the interface between sedimentary structures and sedimentology, biogenic structures and paleontology, metamorphic fabric and metamorphic geology, and regional structures (for example, terrane boundaries, subduction zones, transform faults) and tectonics. Initial effort thus focused on the ontology of fault and fold concepts. Each concept has an associated information schema that identifies class (depicting concept) hierarchies, class object properties, and relations linking the classes (along with value ranges and cardinalities). Formalization of a class for computation requires development of definitions that allow unambiguous identification of an observed individual phenomenon as an instance of the class. The formalization is necessary to determine the appropriate information schema to use in describing the concept and the sort of analytical operations that are appropriate for the instance. Based on the North American Geologic Map Data Model—Conceptual Model 1.0 (NADM-C1), a fold was tentatively defined as "one or more curved layers, surfaces, or lines in a rock body [adopted from Davis, 1984, and Hansen, 1971], unified by continuous axial surface, and bounded by inflection points on limbs" (North American Geologic Map Data Model (NADM) Steering Committee Data Model Design Team, 2004). A fault was defined as "a discrete surface separating two rock masses across which one mass has slid past the other [based on Jackson, 1997]". The fault definition is narrower than that given by Jackson and Bates (1997) in that it includes only a single surface. Object properties identified for folds include orientation, wavelength, span, amplitude, symmetry, profile type, shape, interlimb angle, fold model, and age. Related concepts identified include fold train, folded layer, folded surface, axial surface, hinge line, hinge line segment, and axis. Properties identified for faults include age, orientation, slip, slip rate, recurrence interval, surface character, and environment. Related concepts include material substrate, related structures (minor fault, fold, or fracture), segmentation, mineralization, heat production, landform, sedimentary basin, fault process, global positioning system (GPS) velocity field, strain field, fault zone, and fault system. Development of a geoinformatic infrastructure requires much more work to achieve a community consensus on the definition of these basic concepts and their formal representation in a computable information schema. This consensus is probably best developed by integrating both a bottom-up approach, which facilitates individual project-level knowledge representation and innovation, and a top-down approach, which establishes a top-level unifying framework and facilitates knowledge reuse by other interested communities.

References Cited

Davis, G.H., 1984, Structural geology of rocks and regions: New York, N.Y., John Wiley and Sons, 492 p.

Hansen, Edward, 1971, Strain facies: New York, N.Y., Springer-Verlag, 207 p.

Jackson, J.A., and Bates, R.L., eds., 1997, Glossary of geology, 4th ed.: Alexandria, Va., American Geological Institute, 769 p.

North American Geologic Map Data Model Steering Committee, 2004, NADM Conceptual Model 1.0—A conceptual model for geologic map information: U.S. Geological Survey Open-File Report 2004-1334, 58 p., available only online at http://pubs.usgs.gov/of/2004/1334. (Accessed September 11, 2006.) Also published as Geological Survey of Canada Open File 4737, 1 CD-ROM.

Building and Managing a Waveform Data Set for Finite-Frequency Body-Wave Tomography

By Karin Sigloch[1]

[1]Geosciences Department, Princeton University, Princeton, N.J.

We are assembling and managing a large database of seismic waveform data for use in finite-frequency tomography. This new kind of tomographic modeling remedies many shortcomings of the conventional ray-theoretical approach by explicitly taking into account the finite wavelengths of seismic waves. Finite-frequency observables (travel times, amplitudes) require a more sophisticated measurement effort than conventional "phase picks." Measurements are based on cross-correlations of observed and predicted broadband seismograms. The technical challenge is to remove the strong but a priori unknown earthquake signatures so that the weak signals of mantle anomalies may emerge. We present a set of flexible processing, filtering, and visualization tools that allow for quick, intuitive, and extensive quality control. This is crucial for the efficient processing of hundreds of thousands of seismograms.

Towards an Ontology for Volcanoes

By A. Krishna Sinha[1], Grant Heiken[2], Calvin Barnes[3], Ken Wohletz[4], Dina Venezky[5], Peter Fox[6], Deborah L. McGuinness[7], Robert Raskin[8], and Kai Lin[9]

[1]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Va.

[2]Freeland, Wash.

[3]Department of Geology, Texas Tech, Lubbock, Tex.

[4]Los Alamos National Laboratory, Los Alamos, N.M.

[5]U.S. Geological Survey, Menlo Park, Calif.

[6]High Altitude Observatory/Earth and Sun Systems Laboratory/National Center for Atmospheric Research, Boulder, Colo.

[7]Knowledge Systems, Artificial Intelligence Laboratory/Stanford University, and McGuinness Associates, Stanford, Calif.

[8]Jet Propulsion Laboratory, Pasadena, Calif.

[9]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

The near- and far-field effects of volcanic activity are common geologic phenomena observed in many parts of the world and may include possible catastrophic damage near eruptive centers as well as measurable effects on global climate. A comprehensive categorization of types of volcanic activity and its episodic nature is a challenging scientific goal that requires access to most known types of geologic and sensor data. Ontologies commonly provide the conceptual framework that allows logical association of relationships between many geologic objects and their attributes. At a recent semantically enabled science data integration (SESDI) workshop supported by the National Aeronautics and Space Administration, the first ontologic platform for classification of volcanoes, volcanic activity, and eruption phenomenon was addressed. Conceptual relationships between data and phenomena were developed in the form of concept maps, which serve to document these relationships and are requirements for ontologic access to data and integration tools. This formal encoding of volcano terms is needed to support access to and integration of the data in a manner that leverages term meanings rather than term syntax.

Three high-level classes associated directly with volcanism were identified and defined as “volcanic systems,” “volcanic phenomena,” and “climate.” The concept of “volcanic systems” was linked to both plate tectonic and geologic environments for providing links between eruptive style, magmatic composition, and the location of the volcanic field. Additional concepts related to “volcanic systems” include magma plumbing, eruption environment, and the three-dimensional (3D) geometry of the volcanic field. Deformation, eruption, landslides, biologic activity, atmospheric disturbance, hydrothermal alteration, magma motion, and earthquakes were considered to be subclasses of a higher-level class called “phenomena.” The two classes, “phenomena” and “volcanic systems,” are linked through “materials,” (that is, magma and its products).

This organization of concepts contains associations with all data types used by geologists to monitor active volcanic systems and to study volcanoes preserved in the geologic record. The SESDI project will eventually link the volcanic phenomena to recognized changes in the atmosphere for a more robust understanding of climate change.

Managing the Dynamics of Geographic Information—The Case of Urban Land Use Transformation in St. Louis, Mo.

By Zhanli Sun[1] and Brian Deal[1]

[1]Department of Urban and Regional Planning, University of Illinois—Urbana-Champaign, Champaign, Ill.

Land use change has arrived as an issue of national importance. Communities are beginning to question the impacts that new development will have on their lives, infrastructure, support, and services. Patterns of changing land use have been blamed for increases in traffic congestion and commute times, worsening air and water pollution, loss of farmland and open fields, forests and wetland losses, increased flooding, and property tax inflation. But solutions are difficult; the relationships between this diverse set of problems and the policies that affect land use change are complex and difficult to ascertain.

During the past two decades, spatial analysis tools, such as geographic information systems (GIS) and remote sensing (RS) technologies, have been widely deployed to monitor, analyze, and visualize urban growth phenomena. Maps and satellite images, however, are limited to static displays that portray only the current state of the system, without a way to portray the future. Although GIS-based tools provide useful analytical capabilities and have been widely used to assist urban planners, the static mapping concepts on which they are built are clearly insufficient to study the dynamics of urban growth (Hopkins, 1999).

Very recently, computer-based urban-system-simulation models have been employed to forecast and evaluate land use change (Batty and Xie, 1994; Birkin, 1994; Landis, 1994; Engelen and others, 1995; Wu and Webster, 2000; Waddell, 2002). These models represent a spatial and dynamic approach that enables planners to view and analyze the future outcomes of current decisions and policies before they are put into action. These models have the ability to help improve our fundamental understanding of the dynamics of land use transformation and the complex interactions between urban change and sustainable systems (Deal, 2001). These spatial dynamic modeling techniques are becoming essential elements in the Planning Support System (PSS) literature (Hopkins, 1999; Kammeier, 1999).

Although more essential and relevant, spatial dynamic urban modeling is still in its infancy. Few models have been built that are able to represent the complex dynamics of urban land-use change that are consistent with observable data

(Almeida, 2003). As a result, few such models are operational and can be used to assist in real-world urban planning practices.

We present here a comprehensive dynamic spatial urban simulation model, the Land Use Evolution and Impact Assessment Model (LEAM). LEAM was originally developed as a research project by an interdisciplinary team of researchers at the University of Illinois with support from the National Science Foundation. After a successful full-scale pilot application in Peoria, Ill., LEAM was selected to assist planning practices in the St. Louis metropolitan area, as part of the Department of Defense encroachment analysis, and the Smart Growth initiative introduced by the State of Illinois. Described here is a bi-state application of LEAM consisting of the five counties in southwestern Illinois and the five counties in east-central Missouri that make up the St. Louis metropolitan region, with emphasis on its usefulness in assisting the practice of planning in the region.

References Cited

de Almeida, C.M., Batty, Michael, Monteiro, A.M.V., Camara, Gilberto, Soares-Filho, B.S., Cerqueira, G.C., and Pennachin, C.L., 2003, Stochastic cellular automata modeling of urban land use dynamics; empirical development and estimation: Computers, Environment and Urban Systems, v. 27, p. 481-509.

Batty, M., and Xie, Y., 1994, From cells to cities: Environment and Planning B—Planning & Design, v. 21, p. S31-S38.

Birkin, M., 1994, Urban landscape dynamics—A multilevel innovation process, *in* Montanari A., Curdes, G., and Forsyth, L., eds., Review of urban landscape dynamics: Environment and Planning A, v. 26, p. 1480.

Deal, B., 2001, Ecological urban dynamics—The convergence of spatial modelling and sustainability: Building Research and Information, v. 29, p. 381-393.

Engelen, G., White, R., Uljee, I., and Drazan, P., 1995, Using cellular-automata for integrated modeling of socio-environmental systems: Environmental Monitoring and Assessment, v. 34, p. 203-214.

Hopkins, L.D., 1999, Structure of a planning support system for urban development: Environment and Planning B—Planning & Design, v. 26, p. 333-343.

Kammeier, H.D., 1999, New tools for spatial analysis and planning as components of an incremental planning-support system: Environment and Planning B—Planning & Design, v. 26, p. 365-380.

Landis, J.D., 1994, The California urban futures model—A new generation of metropolitan simulation models: Environment and Planning B—Planning & Design, v. 21, p. 399-420.

Waddell, P., 2002, UrbanSim—Modeling urban development for land use, transportation, and environmental planning: Journal of the American Planning Association, v. 68, p. 297-314.

Wu, F.L., and Webster, C.J., 2000, Simulating artificial cities in a GIS environment; urban growth under alternative regulation regimes: International Journal of Geographical Information Science, v. 14, p. 625-648.

Toward a Better Understanding of Failures in Grid Systems

By Luca Telloli[1], Flavio Junqueira[2], Keith Marzullo[2], Sandeep Chandra[3], and Karan Bhatia[3]

[1]Department of Computer Science, University of Bologna, Bologna, Italy

[2]Department of Computer Science and Engineering, University of California—San Diego, La Jolla, Calif.

[3]San Diego Supercomputer Center, University of California—San Diego, Calif., La Jolla, Calif.

Scientific applications, such as large-scale simulations, require a significant amount of computational resources and storage capabilities. Additionally, scientists in different locations wish to access and share data and results of computations over heterogeneous networks, such as the Internet. A solution to these demands is the interconnection of a number of cooperating yet independent computational infrastructures into a single system, commonly known as a “Grid.”

The large-scale nature of Grids exposes them to all kind of failures that can affect the system in different ways—crashes can cause unavailability of a node, network problems can disconnect multiple nodes, and failures of storage servers can cause processes to stall. In addition, the geographically dispersed nature of such systems makes it harder to handle failures, as no single person has access to all the resources simultaneously. The ability to provide service despite failures is, therefore, a crucial issue.

The first step towards this goal is to understand the nature of failures in wide-area systems. In September 2005, the San Diego Supercomputer Center and the Department of Computer Science and Engineering at the University of California—San Diego started a collaborative project on Failures In Grid Systems (FIGS). The project consists of collecting and analyzing failure data from the Geosciences Network (GEON) grid infrastructure. The main goals of the project are as follows: (1) classify common failures, (2) study the dependency of failures, (3) build realistic failures models, and (4) validate theoretical models with actual data analysis.

Our methodology for collecting data is different from previous works in measurement—we collect snapshots periodically for each node and automatically parse and analyze them to extract information on failures. This procedure allows for a fine-grained classification that includes partitions, reboots, and crashes.

This poster illustrates data collected for a period of eight weeks; sufficient material for a deeper analysis is expected in a few more months. Figure 1 illustrates partitions and their duration during the month of February 2006, when six out of eight monitored clusters experienced some kind of failure, with over 30 partitions and a few crashes.

A preliminary analysis shows that most of the causes of unavailability are short network partitions and that the mean time to repair for crashes is significantly longer than the time for a partition to heal.

With the data that we will have in the future, we expect to confirm the trends we observed and to produce failure models that capture accurately the behavior of the system and yet are useful for the design of algorithms to achieve higher availability in such systems as GEON. These failure models will enable solutions to common problems on Grid systems, such as data replication and consistency.

An ebRIM-Based Catalogue Service Information Model for Virtual Geospatial Data

By Yaxing Wei[1], Aijun Chen[1], Yuqi Bai[1], and Liping Di[1]

[1]Center for Spatial Information Science and Systems, George Mason University, Greenbelt, Md.

At present, many groups around the world collect tremendous amounts of geospatial data from dozens of satellites inspecting the Earth, and the amount of data continues to increase rapidly. These data can be used in many areas, such as in agriculture, flood monitoring, and coastal-zone monitoring. The raw, low-level geospatial data that are collected, however, hide much useful information needed by end users and, therefore, cannot be used directly. High-level, customized data based on user requirements are the most useful to end users. The user requirements in different areas are diverse and change constantly, so it is unreasonable to generate the high-level data in advance. It is reasonable, however, to produce high-level data on the fly when users need them. We propose a new concept—virtual geospatial data—to represent this kind of high-level data. Virtual geospatial data does not really exist but can be generated on the fly based on user requirements when requested. Users can customize the virtual data by specifying necessary attributes and composing an abstract model with predefined service types and data types to represent how the virtual data can be generated from existing data by chaining services.

In a geospatial information system, especially the one supporting virtual geospatial data, the catalogue service plays a very important role. Here, based on the ebXML Registry Information Model (ebRIM) version 2.5, we propose an extended catalogue-service information model that supports the modeling, archiving, managing, querying, and reusing of virtual geospatial data. We use the standard processes ebRIM provides to extend the information model to support virtual geospatial data. The extension focuses on the following four aspects: (1) New classification schemes for data types and service types are defined. Those predefined data types and service types are used as the components to model virtual geospatial data. (2) The service class in the ebRIM is extended using slots to fully describe a Web service instance, such as its operations, input and output parameters, and default values for some input parameters. (3) A new class, DataGranule, is created that represents both real and virtual geospatial data by extending the RegistryObject class in ebRIM with elements from ISO 19115. Further elements are added to keep the information about the abstract model of virtual geospatial data. (4) The ebRIM is extended to manage the relationships among the objects in the catalogue.

The extended catalogue-service information model is suitable for the management of data types, service types, geospatial data, geospatial Web services, and their relationships. It supports those applications in a geospatial information system to model, archive, acquire, and produce virtual geospatial data. We have successfully applied the information model in our two projects, each funded by the National Aeronautics and Space Administration—"Integration of OGC and Grid Technologies for Earth Science Modeling and Applications" and "Geobrain."

Online Geologic Map Information Meets Institutional Memory

By Gerald Weisenfluh[1]

[1]Kentucky Geological Survey, Lexington, Ky.

The Kentucky Geological Survey is developing an Internet map service to provide users with seamless 1:24,000-scale geologic maps along with almost all the other kinds of data that the Survey maintains. The map service Web site at http://kgsmap.uky.edu/website/KGSGeology/viewer.asp uses Environmental Systems Research Institute, Inc.'s (ESRI's) ActiveX connector to display maps, whereas most information is stored in relational databases and retrieved with Active Server Pages (ASP) database queries. The interface allows users to customize the geologic map by specifying which geology themes to display; providing alternative classifications of the rock units, such as primary lithology or karst potential; and adding additional data themes, such as sinkholes or oil wells. Database functions are included to provide access to descriptive information derived from the geologic maps or other sources.

The underlying data model for geologic information has a central component of geographic extent. Most geologic descriptions, including observations, publications, images, or anecdotes, can be assigned a map footprint that facilitates associating the information with a user's area of interest. In the course of the project, tools have been programmed to allow geologists to submit their photographic and textual evidence, observations, and interpretations to the database. These submissions will be accessible from the Internet map, and identified by source and scale of observation.

Metadata Integrated Data Analysis Server (MIDAS)

Ruixin Yang[1], Yujie Zhao[1], and Menas Kafatos[1]

[1]Center for Earth Observing and Space Research, George Mason University, Fairfax, Va.

The Distributed Metadata Server (DIMES) was developed for metadata search and especially for metadata navigation based on the Extensible Markup Language (XML) technology. DIMES is designed to be flexible yet simple, with consideration of small data providers. DIMES uses XML to represent, store, retrieve, and interoperate metadata in a distributed environment. In addition to regular metadata searching, DIMES provides a Web-based metadata navigation interface by using the "nearest neighbor search." A grid analysis and display system (GrADS) data server (GDS) enhances the Open Source Project for a Network Data Access Protocol (OpeNDAP) by tightly integrating OpeNDAP and GrADS to enable the application of the power of GrADS analysis over the Internet. GDS enables on-the-fly, server-side data analysis and manipulation and, therefore, reduces network traffic for large amounts of data or data from multiple sources. In the next stage of development, DIMES (the metadata server) and GDS (the data server) are integrated to form a metadata integrated data analysis server (MIDAS) for interactive access to both metadata and data. The power of MIDAS is reflected in the consistency between the data server and the metadata server and in the close coupling between them. Middle layers are added to combine the systems and to help users generate GDS Universal Resource Locator (URL) addresses. This solution may be the only way to provide useful access to the ever-increasing amount of Earth-observing and model data, such as data from the Earth Observation System (EOS), for re-analysis and model output.

The GEON Portal—Scientific Portal Development for Research and Education

By Choonhan Youn[1], Dogan Seber[1], Kai Lin[1], Ashraf Memon[1], and Chaitan Baru[1]

[1]San Diego Supercomputer Center, University of California—San Diego, La Jolla, Calif.

Scientific portals provide an environment that simplifies access to a variety of resources and distributed computational services. In scientific communities, portals become highly desired as scientific research moves toward using interdisciplinary research that requires access to nontraditional datasets and tools. To address the needs of the geosciences community, we are developing a portal for the Geosciences Network (GEON) community. Our efforts are focused on providing several key features that range from user logins, data access mechanisms, and personal workspaces, to analysis environments. Providing an end-to-end scientific analysis environment is a critical need in the geosciences community. The current technology enables us to build user-friendly environments that require very simple hardware and software from the user side to be able to access all that GEON offers. We also address the need to support education and research activities for a field whose computational needs are increasing rapidly and spans a broad range. Based on these scientific challenges, we are building the GEON computation and data portal and its advanced back-end services.

We implement the GEON portal framework in the traditional three-tier architecture based on the service-oriented architecture (SOA) model. These tiers separate functionality and responsibility, allowing us to develop services independently. Users are able to connect to the portal through a Web browser and access Web services through their standalone client applications using the over-the-wire connection protocols at the front end. The middle tier consists of two basic sections: the user interface server running a portlet engine and a distributed Grid/Web-Service-based middle tier. The Web server runs a portlet engine that contains portlet components. These components may implement specific local-service interfaces and integrate a variety of distributed services. The user interface server is responsible for aggregating and customizing the various components into the portal page of the user. The portlet container defines how user interface components can be plugged in and managed by portal administrators and users. For example, GridSphere Portal Framework, which we are using currently, is an open-source portlet container system that provides this capability. Service components consist of a series of Web/Grid services that can provide access to remote back-end services, such as high-performing computers (HPCs) running portable batch system (PBS) or other queuing systems, data storage devices, mapping tools, and visualization toolkits.

Geoinformatics 2006 Participants

- Nadine Alameh, MobiLaps LLC, nadinesa@mobilaps.com
- Kenneth Allender, Science Applications International Corporation, allender@usgs.gov
- Dave Almeida, Hunter College, CARSI Lab, City University of New York, dalmeida@trans-tek.com
- Roddy Amenta, James Madison University, amentarv@jmu.edu
- Eric Anderson, U.S. Geological Survey, ericanderson@usgs.gov
- Yoshihito Arai, JAMSTEC (Japan Agency for Marine-Earth Science and Technology), nolsont@jamstec.org
- Robert Arko, Columbia University, arko@ldeo.columbia.edu
- Luis Armendariz, Computational Infrastructure for Geodynamics, California Institute of Technology, luis@geodynamics.org
- Ramon Arrowsmith, Arizona State University, ramon.arrowsmith@asu.edu
- Hassan Babaie, Georgia State University, hbabaie@gsu.edu
- Judy Back, U.S. Geological Survey, jback@usgs.gov
- Yuqi Bai, George Mason University, ybai1@gmu.edu
- John Bailey, Alaska Volcano Observatory, jbailey@gi.alaska.edu
- Peter Bajcsy, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, pbajcsy@ncsa.uiuc.edu
- Tawan Banchuen, The Pennsylvania State University, txb213@psu.edu
- Margaret Banton, San Diego Supercomputer Center, mbanton@sdsc.edu
- Calvin Barnes, Texas Tech University, Cal.Barnes@ttu.edu
- Chaitan Baru, San Diego Supercomputer Center, baru@sdsc.edu
- Melissa Batum, Minerals Management Service, mbatum@yahoo.com
- Imadeddine Bazzi, George Mason University, imad_ine@hotmail.com
- Jeanne Behnke, National Aeronautics and Space Administration, jeanne.behnke@nasa.gov
- Gregory Bensen, University of Colorado, gbensen@cires.colorado.edu
- Bora Beran, Drexel University, bb63@drexel.edu
- Luis Bermudez, Monterey Bay Aquarium Research Institute, bermudez@mbari.org
- Angelo Bertolli, National Aeronautics and Space Administration, abertolli@innovim.com
- Ronald Blom, Jet Propulsion Lab, ronald.blom@jpl.nasa.gov
- Arnold Boedihardjo, Virginia Polytechnic Institute and State University, mathwise@vt.edu
- Serkan Bozkurt, U.S. Geological Survey, sbozkurt@usgs.gov
- Sally Brady, U.S. Geological Survey, srbrady@usgs.gov
- Jennifer Brennan, National Aeronautics and Space Administration, Jennifer.L.Brennan.1@gsfc.nasa.gov
- Boyan Brodaric, Geological Survey of Canada, brodaric@nrcan.gc.ca
- Debra Buczkowski, Johns Hopkins University, Debra.Buczkowski@jhuapl.edu
- Steven Buczkowski, Howard Community College, steven@inphotoveritas.com
- Scott Callaghan, University of Southern California, scottcal@usc.edu
- Oscar Carrasco, George Mason University, ocarras1@gmu.edu
- Cinzia Cervato, CHRONOS, cinzia@iastate.edu
- Peter Cervelli, U.S. Geological Survey, pcervelli@usgs.gov
- Sandeep Chandra, San Diego Supercomputer Center, chandras@sdsc.edu
- Aijun Chen, George Mason University, achen6@gmu.edu
- Feng Chen, Virginia Polytechnic Institute and State University, chenf@vt.edu
- Jonathan R. Childs, U.S. Geological Survey, jchilds@usgs.gov
- Marge Cole, National Aeronautics and Space Administration, marjorie.c.cole@gsfc.nasa.gov
- Christopher Condit, University of Massachusetts—Amherst, ccondit@geo.umass.edu
- Eric Cowgill, University of California—Davis, cowgill@geology.ucdavis.edu

- Maria Luisa Crawford, Bryn Mawr College, mcrawfor@brynmawr.edu
- Christopher Crosby, Arizona State University, chris.crosby@asu.edu
- Beata Csatho, The Ohio State University, csatho.1@osu.edu
- Upendra Dadi, George Mason University, udadi@gmu.edu
- Jing Dai, Virginia Polytechnic Institute and State University, daij@vt.edu
- James Damario, Harford Community College, jdamario@harford.edu
- Brian Davis, Science Applications International Corporation, bdavis@usgs.gov
- Nicholas Del Rio, University of Texas—El Paso, ndel2@utep.edu
- Meixia Deng, George Mason University, mdeng@gmu.edu
- Branko Djapic, Columbia University, bdjapic@ciesin.columbia.edu
- Rebecca L. Dodge, University of West Georgia, rdodge@westga.edu
- Asli Garagon Dogru, Bogazici University, garagon@boun.edu.tr
- David Donato, U.S. Geological Survey, didonato@usgs.gov
- Raimundo Dos Santos, Virginia Polytechnic Institute and State University, rdossant@vt.edu
- Kelvin Droegemeier, University of Oklahoma, kkd@ou.edu
- John Evans, National Aeronautics and Space Administration, john.evans@gsfc.nasa.gov
- Hong Fan, George Mason University, hfan2@gmu.edu
- Gyorgy Fekete, The Johns Hopkins University, gfekete@pha.jhu.edu
- Douglas Fils, CHRONOS, fils@chronos.org
- John Findley, Booz Allen Hamilton, j_findley@verizon.net
- Adam Forte, University of California—Davis, forte@geology.ucdavis.edu
- Ian Foster, Argonne National Laboratory, foster@mcs.anl.gov
- Donna Foulke, U.S. Geological Survey, donna_foulke@usgs.gov
- Peter Fox, National Center for Atmospheric Research, pfox@ucar.edu
- Jean Freeney, U.S. Geological Survey, jfreeney@usgs.gov
- Anne Frondorf, U.S. Geological Survey, anne_frondorf@usgs.gov
- Mark Gahegan, The Pennsylvania State University, mng1@psu.edu
- Kevin Gallagher, U.S. Geological Survey, kgallagher@usgs.gov
- Irbis Gallegos, University of Texas—El Paso, irbisg@utep.edu
- Yongli Gao, East Tennessee State University, gaoy@etsu.edu
- Martha Garcia, U.S. Geological Survey, mgarcia@usgs.gov
- Anthony Gary, University of Utah, tgary@egi.utah.edu
- Ann Gates, University of Texas—El Paso, agates@utep.edu
- Ryan Gerard, National Aeronautics and Space Administration, ryan.gerard@gsfc.nasa.gov
- Art Goldstein, National Science Foundation, agoldste@nsf.gov
- David Govoni, U.S. Geological Survey, dgovoni@usgs.gov
- Jim Graham, Colorado State University, jim@nrel.colostate.edu
- John Graybeal, Monterey Bay Aquarium Research Institute, graybeal@mbari.org
- Linda Gundersen, U.S. Geological Survey, lgundersen@usgs.gov
- Thomas Gunther, U.S. Geological Survey, thomas_gunther@usgs.gov
- Jafar Hadizadeh, University of Louisville, hadizadeh@louisville.edu
- Bethany Hall, Middle Tennessee State University, bhall1821@hotmail.com
- Bruce Hemingway, U.S. Geological Survey, bhemingw@usgs.gov
- Christel Hennet, EarthScope, chennet@earthscope.org
- J.C. Herz, National Science Foundation, jnhq@yahoo.com
- Derek Holom, University of Idaho, derekholom@gmail.com
- Hanjo Jeong, George Mason University, hjeong@gmu.edu
- Ying Jin, Virginia Polytechnic Institute and State University, jiny@vt.edu

- Leonard Johnson, National Science Foundation, lejohnson@nsf.gov
- Bruce Johnson, U.S. Geological Survey, bjohnson@usgs.gov
- Salah Jubeh, MAICH, s_jubeh@yahoo.com
- Randy Keller, University of Texas—El Paso, keller@utep.edu
- Larry Kerschberg, George Mason University, kersch@gmu.edu
- Karen Klima, U.S. Geological Survey, kklima@usgs.gov
- Jens Klump, GeoForschungsZentrum Potsdam, jens.klump@gfz-potsdam.de
- Rama Kotra, U.S. Geological Survey, rkotra@usgs.gov
- Yufeng Kou, Virginia Polytechnic Institute and State University, ykou@vt.edu
- Vladik Kreinovich, University of Texas—El Paso, vladik@utep.edu
- Robert Kucks, U.S. Geological Survey, rkucks@usgs.gov
- Mohi Kumar, American Geophysical Union, mkumar@agu.org
- John LaBrecque, National Aeronautics and Space Administration, John.LaBrecque@nasa.gov
- David Lambert, National Science Foundation, dlambert@nsf.gov
- Walter Landry, Caifornia Institute of Technology, walter@geodynamics.org
- Philip Larkin, National Aeronautics and Space Administration, philip.larkin@nasa.gov
- Bih-Ru Lea, University of Missouri—Rolla, leabi@umr.edu
- Brandon League, University of Tennessee, bleague@utk.edu
- Kerstin Lehnert, Lamont-Doherty Earth Observatory of Columbia University, lehnert@ldeo.columbia.edu
- Margaret Leinen, National Science Foundation, mleinen@nsf.gov
- Zhang Lijuan, China Institute of Geotechnical Engineering, philiplzy@yahoo.com.cn
- Kai Lin, San Diego Supercomputer Center, klin@sdsc.edu
- Mian Liu, University of Missouri—Columbia, lium@missouri.edu
- Ben Logan, Virginia Polytechnic Institute and State University, blogan@vt.edu
- Qifeng Lu, Virginia Polytechnic Institute and State University, lqf74cn@yahoo.com
- Chang-Tien Lu, Virginia Polytechnic Institute and State University, ctlu@vt.edu
- Bertram Ludaescher, University of California—Davis, ludaesch@ucdavis.edu
- Peter Lyttle, U.S. Geological Survey, plyttle@usgs.gov
- Philip Maechling, Southern California Earthquake Center, maechlin@usc.edu
- Robert Matthias, U.S. Geological Survey, rmatthias@usgs.gov
- Deborah McGuinness, Stanford University and McGuinness Associates, dlm@ksl.stanford.edu
- Charles Meertens, University NAVSTAR Consortium, meertens@unavco.org
- Ashraf Memon, San Diego Supercomputer Center, amemon@sdsc.edu
- Ghulam Memon, San Diego Supercomputer Center, gmemon@sdsc.edu
- William Miller, U.S. Geological Survey, bmiller@usgs.gov
- Douglas Miller, The Pennsylvania State University, miller@eesi.psu.edu
- Wang Mingzhu, China Institute of Geotechnical Engineering, philiplzy@yahoo.com.cn
- Drew Mirante, Bryn Mawr College, dmirante@brynmawr.edu
- Prasenjit Mitra, The Pennsylvania State University, pmitra@ist.psu.edu
- Joanna Muench, IRIS Data Management Center, joanna@iris.washington.edu
- Richard Murnane, Bermuda Biological Station for Research, rmurnane@bbsr.edu
- Ullas Nambiar, University of California—Davis, ubnambiar@cs.ucdavis.edu
- Viswanath Nandigam, San Diego Supercomputer Center, viswanat@sdsc.edu
- Pragyansmita Nayak, George Mason University, pnayak@gmu.edu
- Mark Nell, University of Idaho, nell7735@uidaho.edu

- Francisca Oboh-Ikuenobe, University of Missouri—Rolla, ikuenobe@umr.edu
- Jai Ho Oh, Pukyong National University, jhoh@pknu.ac.kr
- Emmanuel Gbenga Olagunju, Linkoping University, emagolagunju@yahoo.co.uk
- John Oldow, University of Idaho, oldow@uidaho.edu
- Annette Olson, U.S. Geological Survey, alolson@usgs.gov
- Gilbert Ornelas, University of Texas—El Paso, gtornelas@utep.edu
- Hardik Patel, Virginia Polytechnic Institute and State University, hkpatel@vt.edu
- Richard Pearlson, RAPSAT, rapsat@care2.com
- Vincent Pellerito, Delaware Geological Survey, vpelleri@udel.edu
- David Percy, Portland State University, percyd@pdx.edu
- Fred Pieper, Institute for the Application of Geospatial Technology, fpieper@iagt.org
- Frances Pierce, U.S. Geological Survey, fpierce@usgs.gov
- Paulo Pinheiro da Silva, University of Texas—El Paso, paulo@utep.edu
- Emil Platon, University of Utah, eplaton@egi.utah.edu
- Rahul Ramachandran, University of Alabama—Huntsville, rramachandran@itsc.uah.edu
- Robert Raskin, Jet Propulsion Lab, raskin@jpl.nasa.gov
- Joshua Reed, Iowa State University, jareed@iastate.edu
- Allister Rees, University of Arizona, rees@geo.arizona.edu
- Stephen Richard, Arizona Geological Survey and U.S. Geological Survey, steve.richard@azgs.az.gov
- Sarah Roberts, James Madison University, robertsm@jmu.edu
- Salvatore Rodano, Harford Community College, srodano@harford.edu
- Leslie Ruppert, U.S. Geological Survey, lruppert@usgs.gov
- Alok Sahoo, George Mason University, asahoo@gmu.edu
- Leonardo Salayandia, University of Texas—El Paso, leonardo@utep.edu
- Toni Schenk, The Ohio State University, schenk.2@osu.edu
- Jonathan Sears, American Geophysical Union, jsears@agu.org
- Dogan Seber, San Diego Supercomputer Center, seber@sdsc.edu
- Kaye Shedlock, National Science Foundation, kshedloc@nsf.gov
- Xuan Shi, West Virginia University, xshi@geo.wvu.edu
- Karin Sigloch, Princeton University, sigloch@princeton.edu
- A. Krishna Sinha, Virginia Polytechnic Institute and State University, pitlab@vt.edu
- Walter Snyder, Boise State University, wsnyder@boisestate.edu
- Cassie Soeffing, National Aeronautics and Space Administration, cassandra.k.soeffing@nasa.gov
- David Soller, U.S. Geological Survey, drsoller@usgs.gov
- Nancy Stamm, U.S. Geological Survey, nstamm@usgs.gov
- Alan Stevens, U.S. Geological Survey, asteve1@usgs.gov
- Tyler Stevens, National Aeronautics and Space Administration, tstevens@gcmd.nasa.gov
- Zhanli Sun, University of Illinois—Urbana-Champaign, sun2@uiuc.edu
- Luca Telloli, University of California—San Diego, ltelloli@ucsd.edu
- June Thormodsgard, U.S. Geological Survey, thor@usgs.gov
- Richard Troy, Science Tools, rtroy@sciencetools.com
- E. Lynn Usery, U.S. Geological Survey, usery@usgs.gov
- David Valentine, San Diego Supercomputer Center, valentin@sdsc.edu
- Gregory van der Vink, EarthScope, gvdv@earthscope.org
- Dina Venezky, U.S. Geological Survey, dvenezky@usgs.gov
- Arwen Vidal, University of Colorado, arwen.vidal@colorado.edu
- Chris Weaver, The Pennsylvania State University, cweaver@psu.edu
- Stephen D. Weaver, The Pennsylvania State University, sdw135@psu.edu
- Yaxing Wei, George Mason University, ywei@gmu.edu
- Gerald Weisenfluh, Kentucky Geological Survey, jerryw@uky.edu
- Robert Wertz, U.S. Geological Survey, rwertz@usgs.gov

- Richard Wertz, Foundation for Earth Science, dick.wertz@earthsciencefoundation.org
- Nancy Wiegand, University of Wisconsin—Madison, wiegand@cs.wisc.edu
- Michael Williams, University of Massachusetts, mlw@geo.umass.edu
- Michael Wright, University Corporation for Atmospheric Research, mwright@ucar.edu
- Li Wu, George Mason University, lwu1@gmu.edu
- Douglas Yager, U.S. Geological Survey, dyager@usgs.gov
- Ruixin Yang, George Mason University, ryang@gmu.edu
- Xiaoping Yang, SAIC, xiaoping.yang@saic.com
- Phil Yang, National Aeronautics and Space Administration, cyang3@gmu.edu
- Li Yaoqin, China Institute of Geotechnical Engineering, philiplzy@yahoo.com.cn
- Taehun Yoon, The Ohio State University, yoon.122@osu.edu
- Choonhan Youn, San Diego Supercomputer Center, cyoun@sdsc.edu
- Genong Yu, George Mason University., gyu@gmu.edu
- Peng Yue, George Mason University, pyue@gmu.edu
- Karen Zanter, U.S. Geological Survey, kzanter@usgs.gov
- Ilya Zaslavsky, San Diego Supercomputer Center, zaslavsk@sdsc.edu
- Peisheng Zhao, George Mason University, pzhao@gmu.edu
- Sarah Zielinski, American Geophysical Union, szielinski@agu.org
- Herman Zimmerman, National Science Foundation (retired), zimmermanhv@starpower.net

Manuscript approved for publication August 7, 2006.

Prepared by Reston Publications Service Center.

Editing by Marilyn A. Billone, Iris M. Collies, Anna N. Glover, Jeanette C. Ishee, and Revondra McQueen.

Photocomposition and design by Elizabeth D. Koozmin.

For more information concerning this report, please contact Shailaja R. Brady, U.S. Geological Survey, 911 National Center, Reston, VA 20192, srbrady@usgs.gov.

USGS

Brady and others—**Geoinformatics 2006—Abstracts**—Scientific Investigations Report 2006–5201

ISBN 1-411-31129-9

9 781411 311299

Printed on recycled paper