

GEOINFORMATICS – WHAT WE NEED TO DO TO ACCELERATE WEB ACCESSIBILITY

M. Lee Allison¹ and Linda Gundersen²

¹Arizona Geological Survey, 416 W. Congress, #100, Tucson, AZ 85701, USA, lee.allison@azgs.az.gov, ² U.S. Geological Survey, National Center, Reston, VA 20192 USA, lgundersen@usgs.gov

1. THE PROMISE OF GEOINFORMATICS

The promise of geoinformatics as a “distributed, integrated digital information system and working environment that provides innovative means for the study of the Earth-Sun system” has yet to be fulfilled (Fox et al., 2006). Although numerous individual databases, web-accessible tools, and small information systems have been created and more are under development, the Earth science community lacks the ability to integrate data across these disparate systems in an efficient and comprehensive way (Walker, 2007). It has become clear that the community requires an integrated cyberinfrastructure and that a coordinated geoinformatics effort is needed to accomplish it. Researchers recognize that they cannot take advantage of vast information resources without coordination and collaboration and that this poses a serious obstacle to modeling complex earth systems. Wuchty et al. (2007) underscored this increased demand for interoperability by documenting that, “Teams typically produce more frequently cited research than individuals do and this advantage has been increasing over time. Teams now also produce the exceptionally high-impact research....” They conclude that “the process of knowledge creation has fundamentally changed.” To respond to and support that change, it is imperative that the Earth science community develop mechanisms for easier discovery and integration of the wide variety of data critical for achieving the capabilities and synergies offered by a large-scale cyberinfrastructure.

Many challenges exist for creating such an infrastructure; most are not technical but organizational and cultural in nature. Recent workshops on advancing geoinformatics have recognized and addressed this, and have started focusing on how to achieve cooperation, integration, and community governance of a national and global geoinformatics system. The critical stumbling blocks to creating a wide-reaching geoinformatics component of the cyberinfrastructure for the sciences are: agreements on common standards and protocols; engagement of the vast number of distributed data resources; practices for recognition of and respect for intellectual property; a simple data and resource discovery system (distributed integrated catalogue); mechanisms to encourage development of web service tools for analyses; and business models for sustainability (i.e., continuing maintenance and evolution of information resources).

The “lack of a national civil Earth information strategy” was noted by Gail et al (2007). They call for the nation to “commit to a National Earth-Information Initiative to re-evaluate the national process of collecting and using civil Earth information...and the ability to effectively connect scientific developments to societal uses.”

2. GEOSCIENCE INFORMATION NETWORK

In early 2007, the USA's geological surveys agreed to the development of a Geoscience Information Network (GIN) that is distributed, interoperable, uses open source standards and common protocols, respects and acknowledges data ownership, fosters communities of practice to grow, and develops new web services and clients. Interoperability mechanisms in the GIN involve four modular components:

- Agreement on open-source standards and common protocols through the use of Open Geospatial Consortium (OGC) standards.
- A data exchange model that will start by utilizing the geoscience mark-up language GeoSciML (Richard and CGI IWG, in press; Cox and Richard, 2006), which is an Open Geospatial Consortium GML-based application.
- A prototype data discovery tool or catalogue (“National Digital Catalogue” – NDC) developing under the National Geological and Geophysical Data Preservation Program run by the USGS (DPWG, 2006).
- Data integration and visualization tools developed or planned by a number of independent projects as web service applications.

Geological surveys have unique resources and mission-specific requirements that include the gathering, archiving, and dissemination of data. Together these data represent one of the largest, most extensive long-term information resources on the geology of the United States. Currently, however, these data are available in disparate systems which require time and resources to explore, extract, and reformat. Using modern information technology and a virtual "service oriented architecture" that provides common discovery tools and standards, the surveys and the general science community will benefit in multiple ways. First, online data and other informational products from each survey will be more readily available to the world audience and will be more valuable because they will be interoperable. Second, data and applications from external sources, such as USGS's more than 1,000 databases, catalogues, and inventories will be readily accessed and integratable with each participating survey's own data system. Another 1,000 – 2,000 databases exist in the state geological surveys. Third, a large federated data network will create inestimable opportunities for the broader community, including academia and the private sector, to build applications utilizing this huge data resource, and integrate it with other data.

By demonstrating cooperation for data access and interoperability among the federal and state geological surveys, we may be able to serve as a model for broader cooperation in geoinformatics across the entire Earth science community and linkages to other scientific disciplines, especially those with a geospatial aspect. We intend to coordinate the development of this network with other efforts including the National Science Foundation's "Cyberinfrastructure Vision for 21st Century Discovery," and the emerging academic and international efforts in informatics.

The broader NSF-sponsored workshop in March 2007 examined what direction the geoinformatics community in the United States should be taking in terms of developing a National Geoinformatics System. The final report stated that, "It was clear that developing such a system should involve a partnership between academia (in particular efforts supported by the NSF), government, and industry that should be closely connected to the efforts of the U. S. Geological survey and the state geological surveys that were discussed at a meeting in February of 2007" (Allison et al, 2007b; Keller et al, 2007).

Agreements are developing with OneGeology-Europe (www.onegeology.org), a continental consortium that will make geological spatial data held by the geological surveys of Europe discoverable and accessible. This "community of practice" approach means that we will learn, develop, evolve, and coordinate the building of the network with each other and our partners.

When initiated among the geological surveys, any user may go to a geological survey (or other participating) website or portal, enter a distributed science data catalog, and view available data. Because all these data will use a common mark-up language, the user can select and download needed data and load them into any number of their own applications, including in-house, freeware, and proprietary commercial products. The interface would be seamless and near-instantaneous and the original data source would be credited with the download. The information network will expand as others participate.

3. ROLE OF GEOLOGICAL SURVEYS

Geological surveys have dual requirements to collect, archive, and disseminate data for their stakeholders and customers to use, on one hand, and on the other, to access data held by others that will enable the surveys to better carry out their analytical and research duties. Data accessibility enhances this two-way exchange of information.

The geological surveys are primary geoscience data providers and have mandated responsibilities to collect, organize, and distribute this information to the public. Currently, information assets exist in many databases and in many forms. Similarly, organizations have implemented a wide variety of solutions to manage, process, and support research and data stewardship requirements. Some organizations have integrated their data to provide products to the public, and others have developed accessible Internet Map Services. Because of the large investment in these distributed systems, the emerging service architecture must build on existing systems and use protocols, standards, and services to help integrate the information systems and scientific information.

Geological surveys also contribute to the building of standards of practice and fundamental baseline geologic information such as lexicons, geologic maps, and time scales. These contribute directly to the overall geoinformatics efforts. The breadth and depth of survey-based data are so large that collectively they constitute one of the largest if not the largest data resources in the geosciences, in essence, a national data "backbone."

4. VISION FOR A GLOBAL GEOSCIENCE INFORMATION NETWORK

The surveys agreed to the following principles and activities to be undertaken in the next few years to achieve the vision:

- Develop a coordinated national geoscience information network to access and integrate state survey and USGS information resources (data bases, maps, publications, methods, applications, and data services).
- Function as a “community of practice” in development of geoinformatics and the geoscience network
- Develop prototypes (pilots, test beds) to show proof of concept and determine realistic levels of effort and compare costs and benefits while providing immediate benefits in the form of user services.
- Build the network through an iterative and evolutionary process.
- The basic architecture of the framework should be distributed and leverage existing systems, map services, and data, with local autonomy but using standards to enable interoperability.
- Review and adopt standards and protocols for developing the network (including metadata).
- New and existing systems should communicate with an open source (e.g., Open Geospatial Consortium-based) protocols to promote interoperability.
- Accelerate the development and deployment of a nascent international interchange standard for geological data – GeoSciML - which will enable the sharing and exchange of the data within and beyond the geological community and consider proposing it as a standard to FGDC.
- Recognize there are priority data for which we have mission requirements and inherent partnerships amongst the geological surveys. Review these and adopt service definitions, and protocols as appropriate:
 - Geologic maps, hazard data and maps, topographic data, existing map services
 - Publications and bibliographies
 - Observations and analytical measurements, samples and site information
 - Applications and methods, analytical tools
 - Legacy analog data
 - Resource data and maps (minerals, energy, water etc.)
- Encourage clients and services to be developed and facilitate participation and implementation by others; preferably with low overhead and improving business models and needs.
- Reduce philosophical and cultural barriers that impede system development.
- Adhere to a code of conduct that respects and acknowledges data ownership and the work of others. Respect intellectual property and data provenance, use “branding” in data services to acknowledge data sources. Develop usage measurements and utilize them in clients and web services.
- Develop a database citation format
- Policy – acknowledge that geological surveys need to recognize interoperable, web-enabled information resources as part their mission. Seek partnerships to leverage resources, develop, and implement the vision.

5. NEXT STEPS

A joint AASG-USGS Steering Committee is refining the GIN plan, engaging stakeholders and additional participants, and seeking resources to further plan and implement the network, as part of the emerging National Geoinformatics System.

6. REFERENCES

- Allison, L., Dickenson, T., and Gundersen, L. 2007a. "Role of State Geological Surveys and USGS in a Geological Information System for the Nation: A report to the AASG and USGS on the results of a workshop conducted February 21-22, 2007," <http://www.geoinformatics.info/Resources.html>
- Allison, M. L., Dickinson, T., Gunderson, L. C., Keller, G. R., Maidment, D., and Walker, J. D. convenors, 2007b. "Report from the Workshop on Envisioning a National Geoinformatics System for the United States," 6p, <http://www.geoinformatics.info/Resources.html>
- Allison, M. L., and Gundersen, L. C. in press. AASG-USGS Plan for a National Geoscience Information Network (expanded abs), Geoinformatics 2007, San Diego CA.
- Data Preservation Working Group (DPWG) of the National Cooperative Geologic Mapping Program Federal Advisory Committee, October 10, 2006, "Implementation Plan for the National Geological and Geophysical Data Preservation Program," U.S. Geological Survey, 26p.
- Fox, P., Gundersen, L.C., Lehnert, K., McGuinness D., Sinha K., and Snyder W. 2006. At the Fall Meeting: "Toward Broad Community Collaboration in Geoinformatics," Eos Trans. AGU, 87(46), 513.
- Gail, W. B., Lane, N. F., and MacCauley, M. K. 2007. "A U.S. Earth Information Strategy?," Imaging Notes, Summer 2007, p45-46.
- Keller, G. R., Maidment, D., Walker, J. D., Allison, L., Gundersen, L. C., and Dickinson, T. in press. "A Community Workshop and Emerging Organization to Support a National Geoinformatics System in the United States," (expanded abstract), Geoinformatics 2007, San Diego, CA.
- National Science Board, September 2005, "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century," National Science Foundation, NSB-05-40, 88p, www.nsf.gov/pubs/2005/nsb0540.start.htm
- National Science Foundation, Cyberinfrastructure Council, March 2007, "Cyberinfrastructure Vision for 21st Century Discovery," NSF 07-28, 57p.
- Richard, S.M., 2006, "Geoscience concept models," in Sinha, K, and Baru, C., Geoinformatics: Geological Society of America Special Paper 397, p. 81-107
- Richard, S.M., and CGI Interoperability Working Group, in press, "GeoSciML—a GML application for geoscience information interchange," in Soller, D.R., ed., Digital Mapping Techniques '06—Workshop Proceedings, U. S. Geological Survey Open-File Report.
- Cox, S.J.D., and Richard, S.M. 2006. "A formal model for the geologic timescale and GSSP, compatible with geospatial information transfer standards," Geosphere, paper #GEOS-00022R
- Walker, J. D., editor, 2007, "Building a National Geoinformatics System, - A Community Workshop: Introduction and Background Information," 129p, www.geoinformatics.info/Resources.html
- Wuchty, S., Jones, B. F., and Uzzi, B. 2007. "The Increasing Dominance of Teams in Production of Knowledge," Science, v316, #5827, p1036-1039, 10.1126/science.1136099, 10 October 2007; www.sciencemag.org/cgi/content/full/1136099/DC1