Final Report

A Workshop on the Role of State Geological Surveys and U.S. Geological Survey in a Geological Information System for the Nation

NSF Award 0723437 to the Association of American State Geologists
March 1, 2007 to February 29, 2008

M. Lee Allison, Principal Investigator
Association of American State Geologists
c/o Arizona Geological Survey
Tucson, Arizona 85701
lee.allison@azgs.az.gov

Tamara L. Dickinson, Principal Investigator
U.S. Geological Survey
12201 Sunrise Valley Drive
Reston, VA 20912

AASG
Association of American State Geologists

USGS
science for a changing world

Arizona Geological Survey
Open-file Report 08-1

A Workshop on the Role of State Geological Surveys and USGS in a Geological Information System for the Nation

A report on the results of a workshop conducted February 21-22, 2007 and subsequent developments

Summary: Attendees representing the Association of American State Geologists and the U.S. Geological Survey met in Denver February 21-22, 2007 to discuss opportunities for making their data accessible and interoperable. They recommended that the USGS and State Geological Surveys work together to create a distributed national Geological Information Network (GIN) of digital data through the use of common standards and protocols, that does not impinge on ownership or control of data, and that builds on existing data systems. This report was prepared by M. Lee Allison and Tamara Dickinson, Principal Investigators, with contributions from Linda C. Gundersen and John C. Steinmetz.

Executive Summary

The two-day workshop held February 21-22, 2007 included representatives of the Association of American State Geologists (AASG) and the U.S. Geological Survey (USGS). The group recommended that the nation's Geological Surveys[1] should develop a national Geological 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.

Geological Surveys have unique resources and mission-specific requirements that include the gathering, archiving, and dissemination of data. The Surveys[2] will benefit in multiple ways from employing this "service-oriented architecture" approach to share digital information. First, online data and other information products from each Survey will be more readily available to the world and will be more valuable because the data will be interoperable. Second, data and applications from sources external to each Survey, such as USGS's more than 1,000 databases, catalogues, and inventories and the estimated 1,000-2,000 databases among the state geological surveys will be readily accessed and integratable with each participating Survey's own data system. Third, a large federated data network will create opportunities for the broader community, including academia and the private sector, to build applications utilizing this data resource, and integrate it with other data. The work of the Geological Surveys will be enhanced by access to these new data and applications.

Our premise is that a national network of geological survey digital data can be a tipping point in creating a transformational advance in geoinformatics. By demonstrating national cooperation for data access and interoperability among the Geological Surveys, we can serve as a model for broader cooperation in geoinformatics across the entire earth science community.

When completed, we envision a scenario where any user may go to a Geological Survey website, enter a distributed science data catalog, (for instance through a simple piece of software served on each Survey's website) and view available geoscience data for a geographic area. The catalog could be viewed on Google maps or a shaded relief backdrop for example. It would include State Geological Survey and U.S. Geological Survey data and show the kinds of data available, a universal resource locator (URL) for

[1] Geological Surveys refers to the State Geological Surveys and the USGS.
[2] Surveys refers to the State Geological Surveys and the USGS.

the data, and the location of the data. Because all these data will use a common mark-up language, the user can immediately select and download the needed data and load them into any number of applications, including in-house, freeware, and proprietary commercial products. The interface would be seamless and instantaneous and the original data source would be credited with the download. The catalog would attract users to individual Survey websites as well as provide a seamless discovery tool.

The attendees considered carefully the unique mission of Geological Surveys to serve the users in their state and the nation, to provide support for decision makers, to account for use of the data and show the success of their programs, to manage their data assets and build upon previous efforts, and to not engage in excessively expensive and laborious activities to achieve results. We think that the concept for a GIN proposed here, adheres to those principles.

The workshop participants identified three priority areas for near-term successes:

1. Geologic maps, topographic maps, geophysical data, and hazards
2. Publications, bibliographies, databases, and repositories
3. Site observations including mineral extraction, geochemical analyses, physical samples, oil and gas extraction, and water

The participants developed an action plan which included presenting the workshop recommendations to the AASG and USGS for adoption; presenting the results to the larger geoinformatics and geoscience communities; forming a governance structure to implement the plan if adopted by AASG and USGS; selecting test beds for pilot projects; and preparing one or more funding proposals for the pilot projects. It was recommended that initial test beds be conducted for geologic maps and a site observation catalog.

Subsequent to the workshop:

- USGS and AASG formally adopted the recommendations
- GIN may serve as the implementation model for the USGS Data Integration Blueprint
- A GIN Steering Committee was formed to implement the network
- The National Geoinformatics System community agreed to incorporate GIN into the larger system
- OneGeology agreed to make their global system compatible with GIN
- GIN is collaborating with OneGeology-Europe to develop a continent-wide spatial data network
- The software firm ESRI agreed to develop a Geology Data Model for ArcGIS software compatible with GIN
- The National Geological & Geophysical Data Preservation Program is developing a National Data Catalog and inventory of data holdings in Department of Interior (DOI) Bureaus and State Geological Surveys that will become an integral component of GIN

Metrics of workshop results (details are listed at the end of this report):

- 8 publications
- 6 formal presentations at national meetings and workshops
- 6 briefings government, professional, and industry groups
- 6 funding proposals submitted (5 to NSF, 1 to USGS)

Introduction

Data access and interoperability have been discussed in workshops and conferences on geoinformatics over the past five years. However, the topic of the role of State Geological Surveys and the US Geological Survey in a geological information system has been addressed in only a limited way for a specific data set. The Data Preservation Working Group of the National Cooperative Geologic Mapping Program Federal Advisory Committee recently released an "Implementation Plan for the National Geological and Geophysical Data Preservation Program" (Data Preservation Working Group, 2006). The successful development of this limited data partnership leads us to believe that it can be expanded into a comprehensive national geological survey information system.

The Growing Need for Data Accessibility

At the "Geoinformatics 2006" meeting in May 2006, a dominant theme was the challenge of getting people to share raw data. John LaBrecque, head of NASA's Earth Surface and Interior program, said that organizations need to take a federated approach to data management and that different agencies should work together to develop data standards, common nomenclature, and new technologies, to enlist the enthusiasm of the geoscience community (Kumar, 2006). In a recent post to the geoinformatics web public forum, Walt Snyder (Boise State U.) and Kerstin Lehnert (LDEO) made the point that, "What hinders science and public policy decision making, is the lack of complete access to all relevant data and metadata" (http://www.geoinformatics.info/forum/viewtopic.php?t=19).

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.

History

The World Wide Web (WWW) grew out of the combination of URLs, hypertext transfer protocols (HTTP), and hypertext mark-up language (HTML) as a data exchange format, and access to the Internet (Gillies and Cailliau, 2000). It is a system designed and built initially by combining existing components. Subsequent development came from users creating application and implementation software and processes to enhance the system. By adopting these few simple protocols, users have made the Web one of the most powerful tools and significant technological advances in history. Can adoption of a few data access and interoperability protocols by the nation's Geological Surveys provide a similar stimulus to help foster geoinformatics as a significant resource and tool for the earth sciences?

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.

The USGS National Geologic Map Database Project (NGMDB) has been promoting collaborative development of a geologic map database as mandated by the National Geologic Mapping act of 1992 (with reauthorizations in 1997 and 1999). Currently, the project is maintaining a national geologic map catalog (http://ngmdb.usgs.gov/ngmdb/ngm_catalog.ora.html), the Lexicon of geologic names (http://ngmdb.usgs.gov/Geolex/geolex_home.html), and has been a principal sponsor for the development of the North American Geologic Map data model (NADMC1, see NADMSC, 2004), which provides a conceptual foundation for interoperable geologic information services. The project has sponsored the Federal Geographic Data Committee (FGDC) standard for geologic map symbols, is developing a web portal for a distributed network of geologic map data, and has held 10 Digital Mapping Techniques workshops, where a significant body of technical information on State and Federal database-design and standards development is documented.

Another effort of note among the Geological Surveys is the recently completed *"Implementation Plan for the National Geological and Geophysical Data Preservation Program, Data Preservation Report*". This USGS report was created collaboratively with members from the state geological surveys, industry, and academia and submitted to Congress by the Secretary of the Department of Interior in October 2006. Some of the provisions in the plan are relevant to the issue of digital data and coordination will need to occur among the activities proposed for a national GIN and in the data preservation plan. Notable excerpts from the data preservation plan are as follows:

> *As part of the Energy Policy Act of 2005, the USGS is tasked with creating a National Geological and Geophysical Data Preservation Program. This Program will: (1) archive geologic, geophysical, and engineering data, maps, well logs, and samples; (2) provide a national catalog of such archival material; and (3) provide technical and financial assistance related to the archival material. The Program is authorized at $30M per year for five years but no funds have yet been appropriated.*
>
> *The Program is envisioned as a national network of cooperating geoscience materials and data repositories that are operated independently yet guided by common standards, procedures, and protocols for metadata. The holdings of all collections will be widely accessible through a single, common, and mirrored Internet-based catalog, the National Digital Catalog, thus maximizing the availability of and interconnectedness of all the collections.*

The National Geological and Geophysical Data Preservation Program (NGGDPP) focuses on samples and associated data preserved at DOI Bureaus and State Geological Surveys. A cornerstone of the Program is a National Digital Catalog which can be integrated into a system of geoinformatics directory services as part of the information discovery architecture.

<u>Cooperation Across all Communities in Geoinformatics</u>

There is a growing sense that geoinformatics has matured and evolved across the earth science community so that its potential can be more easily achieved. Projects that had long been working independently are seeing the mechanisms and benefits of integration and interoperability. Further, the earth sciences are becoming more familiar and comfortable with information technology, leading to fuller integration of the two fields across a wide spectrum of activities.

There is also a range of technical capabilities across the earth sciences rather than a simple digital divide and the community realizes the need for a system that accommodates the full breadth of technical capabilities. To realize an operational geoinformatics system, more research is needed on interoperability of data, innovative search and retrieval engines, web services, and creation of science applications to use data. Overall, cooperation in the community is now viewed as essential and desirable, versus the old paradigm of holding onto one's data as proprietary. By demonstrating national cooperation for data access and interoperability among the Geological Surveys, we can serve as a model for broader cooperation in geoinformatics.

Attendees of the AASG-USGS workshop in February strongly supported the idea of "communities of practice". This practice is defined by its creator Etienne Wegner in the following way: "Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly". To create the national GIN and to move geoinformatics forward we must encourage inclusiveness, communication, and a process that is iterative and evolutionary. The attendees recommended the formation of working groups to address opportunities and technical challenges and to share knowledge and applications. They also recommended the use wikis and other online collaborative tools to encourage the cross-discipline sharing of knowledge and applications as cyberinfrastructure for the geosciences is developed.

A Tipping Point for Transformation

The Geological Survey's missions can complement and facilitate development of a national GIN as well as benefit greatly from the result. 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."

The premise is that a national network of Geological Survey data can be a tipping point in creating a transformational advance in geoinformatics. Establishment of organizational and technical protocols to seamlessly access and link information from these various government sources will create inestimable opportunities for the broader community to build applications utilizing this data resource, and integrate it with other data. If the Surveys adopt one or more sets of protocols for data interoperability, they could easily become earth science-wide standards, just as the adoption of HTTP, HTML, and URL's as protocols for the Internet led to the success and near universal use of the World Wide Web.

Providing Input into the National Discussion

The AASG-USGS workshop was held in February 2007 in order to have the results available for a broader workshop on "Envisioning a National Geoinformatics System for the U.S." in March 2007 in Denver. The results of the AASG-USGS workshop were presented to the larger academic community at the March meeting. Prior to the March workshop the results of the AASG-USGS workshop were presented to a meeting of the AASG in Washington, D.C. on March 4, 2007 for endorsement of the recommendations. The AASG group met with the USGS Director and senior USGS management on March 5 and the workshop results were collectively discussed and further action determined.

Description of the February 21-22, 2007 Workshop

Twenty-two individuals attended the workshop, representing nine State Geological Surveys and several USGS programs and disciplines (See Appendix A). The participants included a mix of individuals with a technical background and a working knowledge of the service oriented architectures and experienced managers. Having this mix of technical experts and managers facilitated reaching agreement on a variety of issues and recommendations.

The workshop was dominated by discussion and breakout groups rather than formal presentations (Appendix B). There were a few presentations to provide background material on discussion topics. Each session was facilitated by a representative of a state survey and the USGS and was focused on a series of questions that were developed by the organizers prior to the workshop. A list of acronyms (Appendix C) was compiled for the participants to facilitate discussion.

The main points of discussion centered on forming a distributed national GIN that would be a coordinated framework for accessing and integrating state and USGS data and the roles and responsibility of each organization. Workshop participants quickly reached consensus on creating such a framework and the subsequent discussions focused on the components and content of the system. Using case studies, the participants examined the challenges and opportunities to an integrated or coordinated effort among the state geological surveys and USGS. The participants identified several topical opportunities for near term successes. These opportunities included:

1. Geologic maps, topography, imagery, and spatially represented data such as gravity and magnetics
2. Publications and bibliographies
3. Observations, analyses, and measurements
4. Physical samples
5. Applications and methods, analytical tools, open source software, modeling, catalogs
6. Historical (legacy) data
7. Resource extraction sites (minerals, oil and gas production, mines, quarries)
8. Geologic hazards

With further discussion these were grouped into three priority areas for near term successes:

1. Geologic maps, topographic maps, geophysical data, and hazards
2. Publications, bibliographies, databases, and repositories
3. Site observations including mineral extraction, geochemical analyses, physical samples, oil and gas extraction, and water

These three areas were further discussed during breakout sessions. Each breakout group was asked to address four issues:

1. Define content and establish priorities
2. Define the role of the surveys (contributor, governance, other)
3. What has already been done in this area?
4. What would the interface be (portal, catalog)?

During subsequent breakout sessions, the group identified technical and policy implementation issues associated with each of the priority areas. The technical breakout session also identified areas that were appropriate for test beds.

Final Agreements and Recommendations

During the final session of the workshop the participants agreed to the recommendations presented below. In addition, a clear vision emerged out of the discussions with the goal of facilitating data interoperability, discovery, and accessibility.

> **Vision** – The Nation's Geological Surveys support development of a national geoscience information network that is distributed, interoperable, uses open source standards and common protocols, respects and acknowledges data ownership, and fosters communities of practice to grow and develop new web

services and clients.

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

- Develop a coordinated national GIN 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 GIN
- 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 framework 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 system (including metadata).
- New and existing systems should communicate with an open source (OGC-based) protocol such as the widely adopted geography mark-up language (GML) to promote interoperability.
- Test and consider accepting GeoSciML as a protocol and consider proposing as a standard to Federal Geographic Data Committee (FGDC).
- 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.
- 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 in collaboration with the Geoscience Information Society.
- 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.

Next Steps and Action Items

The workshop participants agreed on a series of next steps, many of which have been accomplished.

1. Write report of workshop.
2. Present workshop report and recommendations to AASG and USGS at Liaison Meeting March 4-5, 2007.

3. Present workshop report and recommendations to NSF on March 6, 2007.
4. Report on AASG-USGS activities at EuroGeoinformatics, Edinburgh, UK, March 8.
5. Presentation at NSF Meeting March 14, 2007.
6. Presentation at the May "Geoinformatics 2007" meeting in San Diego.
7. Presentation at 2007 Annual GSA.
8. Establish a AASG/USGS coordinating committee for Geoinformatics–that includes the needed technical expertise and can serve as a steering committee for future geoinformatics activities (April 2007).
9. Establish technical working groups under coordinating committee to address critical needs.
10. Post report of workshop on www.geoinformatics.info with selected handouts and presentations (by April 2007).
11. AASG/USGS solicit test beds very quickly after acceptance (possible test beds include transborder states and PNW, see below) (by April 2007).
12. Initiate test bed projects in geologic maps, site services, and investigate technology for document searches. Create a funding proposal to NSF and USGS to start the test beds (by May 2007).

Test Beds

The attendees recommended that a steering committee and appropriate technical committees be formed to carry out the first steps in implementing the recommendations of this report and creating the test beds that will examine the technology and feasibility of creating the distributed network and the initial products envisioned.

<u>Science Catalog</u>

The science catalog would contain reference to digital databases held at each Geological Survey and currently served on their websites that contained observations made at a point (lat. long.) or in an area (defined by boundary coordinates). When the catalog is completed, we envision a scenario where any user may go to a Geological Survey website, enter a distributed science data catalog, (for instance through a simple piece of software served on each Survey's website) and view available geoscience data. The catalog could be viewed on any of a variety of backdrop, such as Google maps or a shaded relief backdrop. It would include State Geological Survey and USGS data and show the kinds of data available, a URL for the data, and the location of the data. Because all these data will use a common mark-up language, the user can immediately select and download the needed data and load them into any number of their own applications. The interface would be seamless and near-instantaneous and the original data source would be credited with the download. The catalog would attract users to individual Geological Survey websites as well as provide a seamless discovery tool. Workshop participants recommended that the Steering Committee form a small technical committee with membership from the Geological Surveys to examine current software available, best practices in existing catalogs, and catalogs being developed to recommend an approach and a test bed of a variety of data that already exists and can be found at several geological surveys.

<u>Geologic Maps</u>

The International Union of Geological Sciences (IUGS) through its Commission for the Management and Application of Geoscience Information (CGI) is enabling the global exchange of knowledge about geoscience information and systems. For geologic maps, we intend to emulate the IUGS-CGI Data Model Collaboration, a significant international effort intended to harmonize geoscience web services. In the past two years the IUGS/CGI Working Groups have (1) developed GeoSciML, a standard Geographic Markup Language (GML) application, and (2) initiated two multi-country test beds to evaluate GeoSciML and related Open Geospatial Consortium (OGC) compliant web service technologies. The

International Geological Map **Test Bed 2** for GeoSciML was demonstrated recently (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/TestBed2). The test bed allowed display, query and reformatting of geological map data in real time from US, Canada, Sweden, UK, Australia and France using browser-based client applications served from Canada and France as well as a local desktop client application. Functions demonstrated include: 1) Continuous map portrayal with data from all servers, including attribute query; 2) Reclassification of maps according to age and lithology attributes; 3) Download of complex data structures encoded in GeoSciML.

The proposed test bed should be a multi-state/USGS collaboration such as displaying the geology along the boundary with Mexico in the southwest and a test bed among adjacent states such as the Pacific Northwest. Both of these test beds have funded ongoing mapping activities and may have funds available for conducting this work in the short term. Another alternative would be to solicit test beds competitively. This would have to be done quickly to take advantage of end of year funds.

Subsequent developments

<u>Adoption of GIN by USGS and AASG</u>
On March 2, 2007, USGS Director Dr. Mark Myers was briefed on the Workshop results and he formally committed the USGS to carrying out the recommendations. Linda Gundersen was appointed as the leader for the USGS.

On March 5, 2007, the AASG Executive Committee endorsed the workshop recommendations and appointed workshop co-chair Lee Allison to represent AASG on a joint steering committee with the USGS to develop the GIN.

On March 6, 2007, USGS and AASG leaders held a breakout session at the US Department of Interior in Washington, DC to discuss moving forward with implementation of the GIN. Linda Gundersen and Lee Allison were directed to appoint appropriate members from their respective organizations to a USGS-AASG Steering Committee for GIN as the first step.

<u>National Geoinformatics System workshop</u>
A broader NSF-sponsored workshop held March 14-15, 2007 in Denver examined what direction the geoinformatics community in the United States should be taking in terms of developing a National Geoinformatics System (NGS). The final report (Allison, et al, 2007) stated that, “It was clear that developing such a system should involve a partnership between academia, 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*”

The NGS workshop participants formed a Governance Planning Committee that subsequently became the NGS Coordinating Committee to look at options for organizing the community to effectively build the NGS. Lee Allison and Linda Gundersen were appointed to the Coordinating Committee to ensure smooth interaction with the AASG-USGS GIN development.

<u>USGS Data Integration Blueprint</u>
USGS staff responsible for developing an agency-wide Data Integration Blueprint, indicated after the initial workshop that the GIN approach appeared to be the model they should adopt. They confirmed this informally at the “Geoinformatics 2007” meeting in San Diego in June, 2007.

<u>GIN Steering Committee</u>

Allison and Gundersen appointed members from their respective organizations to the GIN Steering Committee (Appendix D). Creation of at least one subcommittee on technical issues is under discussion. The USGS is expected to name a representative shortly from the Water Resources Discipline to complete participation by all four branches of the USGS on the GIN Steering Committee

OneGeology partnership
Following the AASG-USGS workshop, Dr. Harvey Thorleifson, director of the Minnesota Geological Survey, represented the United States at the OneGeology (www.onegeology.org) workshop in Brighton, England and reported on the GIN plan. Eighty one participants from forty three nations and fifty three national and international bodies met March 12-16, 2007, to discuss how to improve the digital accessibility of global, regional and national geological map data.

The core agreement from the OneGeology meeting, known now as the Brighton Accords, is that *OneGeology is a Geological Survey initiative launched in the International Year of Planet Earth, which will make public and Internet-accessible the best available geological map data worldwide, initially at a scale of about 1:1 million, to better address the needs of society.* It was unanimously adopted by the participants.

The group also agreed to cooperate with GIN by adopting a compatible set of standards and procedures so the US and international networks will be seamlessly interoperable.

OneGeology-Europe
The GIN Steering Committee offered formal support to OneGeology-Europe, a coalition of 30 European organizations, including 21 national geological surveys, in a funding proposal to the European Commission eContent*plus* programme for €3.15 million to build an interoperable spatial data network comparable and compatible with GIN.

The project will accelerate the development and deployment of a nascent international interchange standard for geological data, GeoSciML, enabling the sharing of data within and beyond the geological community.

ESRI ArcGIS Geology Data Model
The GIS software company ESRI, committed to building a Geology Data Model (GDM) compatible with GIN for their ArcGIS products, in part because the potential for interoperability not only within the state and federal geological surveys, but also because it seems likely that the GIN standards will become widespread throughout the earth science community.

Allison gave an informal briefing at a town hall meeting of the petroleum and mineral users groups at the ESRI Users Conference in San Diego in June 2007, where the community expressed overwhelming support to create the GDM as proposed.

National Data Catalog
The USGS-administered National Geological and Geophysical Data Preservation Program (NGGDPP - http://www.usgs.gov/contracts/NGGDP/index.html) is a newly-funded cooperative program with the AASG, as authorized by Congress in the 2005 Energy Policy Act. NGGDPP has committed to developing a National Data Catalog that will be an integral component of GIN. Tamara Dickinson (USGS) and Stephen Richard (AZGS) were both participants in the AASG-USGS GIN workshop and are providing leadership in National Digital Catalog.

Acknowledgements

Much of the body of this report is taken from a preliminary report prepared shortly after the

February 2007 workshop, by M. Lee Allison, Tamara L. Dickinson, and Linda C. Gundersen and posted online at http://www.geoinformatics.info/Resources.html (full reference below).

John C. Steinmetz and Stephen M. Richard contributed to this final report and provided editorial review.

This material is based upon work supported by the National Science Foundation under Grant No. 0723437. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Allison, M. Lee, Tamara L. Dickinson, Linda C. Gundersen, G. Randy Keller, David Maidment, and J. Douglas Walker, convenors, 2007, "Report from the Workshop on Envisioning a National Geoinformatics System for the United States," (http://www.geoinformatics.info/Resources/NSF%20workshop%20on%20National%20Geoinfo%20System%20for%20NSF.pdf

Data Preservation Working Group 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, 28p.

Gillies, James, and Robert Cailliau, "How the Web Was Born: the Story of the World Wide Web," Oxford University Press, 2000, 372p.

Kumar, Mohi, 2006, "Geoinformatics 2006," EOS, American Geophysical Union, vol 87, no. 44, Oct 31, p481.

North American Geologic-Map Data Model Steering Committee (NADMSC), 2004, NADM Conceptual Model 1.0, "A Conceptual Model For Geologic Map Information," U.S. Geological Survey Open-File Report 04-1334, 61p., http://pubs.usgs.gov/of/2004/1334/2004 - 1334.pdf

Walker, J. Douglas, editor, 2007, "Building a National Geoinformatics System - A Community Workshop, Introduction and Background Information," March 14-15, 2007, Denver, CO, 127p., http://www.geoinformatics.info/Resources.

Publications resulting from award

Allison, Lee, Tamara L. Dickenson, and Linda C. Gundersen, 2007, *The 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*, 18p., http://www.geoinformatics.info/Resources.html

Allison, M. Lee, Tamara L. Dickinson, Linda C. Gunderson, G. Randy Keller, David Maidment, and J. Douglas Walker, convenors, 2007, *Report from the Workshop on Envisioning a National Geoinformatics System for the United States*, 6p, http://www.geoinformatics.info/Resources.html

Allison, M. Lee and Linda C. Gundersen, 2008, "How an Integrated State-Federal Geoscience Information Network Can Be Applied in the Borderlands," in: *Proceedings of a USGS Workshop on Facing Tomorrow's Challenges Along the U.S.-Mexico Border—Monitoring, Modeling, and Forecasting Change Within the Arizona-Sonora Transboundary*

Watersheds, 63p, eds, Laura M. Norman, Derrick D. Hirsch, and A. Wesley Ward, USGS Circular 1322, pp 9-10

Keller, G. Randy, David Maidment, J. Douglas Walker, M. Lee Allison, Linda C. Gundersen, and Tamara L. Dickinson, 2007, *A Community Workshop and Emerging Organization to Support a National Geoinformatics System in the United States*, (expanded abstract), pp.75-76, *in* ,S. R. Brady, A.K. Sinha, and L.C. Gundersen, editors, "Geoinformatics 2007-Data to Knowledge," Proceedings: U.S. Geological Survey Scientific Investigations Report 2007-5199, 104p.

Allison, M. Lee, and Linda C. Gundersen, 2007, *Association of American State Geologists (AASG)-USGS Plan for a National Geoscience Information Network* (expanded abs), pp 76-77, *in* S.R. Brady, A.K. Sinha, and L.C. Gundersen, editors, "Geoinformatics 2007-Data to Knowledge," Proceedings: U.S. Geological Survey Scientific Investigations Report 2007-5199, 104p.

Allison, M. Lee and Linda C. Gundersen, 2007, *Geoinformatics – What We Need to do to Accelerate Web Accessibility,* in "Three Dimensional Geologic Mapping for Ground Water Applications," Geological Society of America Workshop Extended Abstracts, October 27, 2007, Denver, CO, Minnesota Geological Survey Open-file Report 07-04.

Allison, M. Lee, Linda C. Gundersen, Tamara Dickinson, and John Steinmetz, 2007, "*Building the Geological Survey Information Network*," Geological Society of America *Abstracts with Programs*, vol. 39, no. 6, p.617.

Allison, M. Lee, Linda C. Gundersen (2007), *Geoscience Information Network*, Eos Trans. AGU*,* vol.*88*(52), Fall Meet. Suppl., Abstract IN21C-05

Presentations resulting from award

Allison, M. Lee, and Linda C. Gundersen, NSF Workshop on Envisioning a National Geoinformatics System for the U.S., March 14, 2007, Denver, CO, *Role of geological surveys in a national information network*

Allison, M. Lee and Linda C. Gundersen, USGS Workshop on "Facing Tomorrow's Challenges along the US-Mexico Border, March 20, 2007, Tucson, AZ, *How an Integrated State-Federal Geoscience Information Network Can be Applied in the Borderlands*

Allison, M. Lee, and Linda C. Gundersen, Geoinformatics 2007, June 2007, San Diego, CA, *Association of American State Geologists (AASG)-USGS Plan for a National Geoscience Information Network*

Keller, G. Randy, David Maidment, J. Douglas Walker, M. Lee Allison, Linda C. Gundersen, and Tamara L. Dickinson, Geoinformatics 2007, June, 2007, San Diego, CA, *A Community Workshop and Emerging Organization to Support a National Geoinformatics System in the United States*,

Allison, M. Lee, Geological Society of America Annual Meeting "Hot Topics" panel -"*Google Earth, Geology Style,*" October 28, 2007, Denver, CO, "Web services for the Geoscience Information Network"

Allison, M. Lee, Linda C. Gundersen, American Geophysical Union Fall Meeting, December, 2007, San Francisco, CA, *Geoscience Information Network*

Scheduled presentations

M. Lee Allison and Linda C. Gundersen, April 23, 2008, "Building a Global Interoperable Geoscience Information Network," AAPG Annual Meeting, San Antonio, TX [poster]

Workshop results briefings

M. Lee Allison, March 4, 2007, AASG Liaison meeting, Jury's Hotel, Washington DC

M. Lee Allison and Linda C. Gundersen, March 5, 2007, co-leaders, geoinformatics workshop breakout session AASG-USGS coordination meeting, US Dept of Interior, Washington, DC

M. Lee Allison, March 6, 2007, NSF-EAR briefing, Arlington, VA

M. Lee Allison, May 15, 2007, Robotic Mining Conference, University of Arizona, Tucson, AZ

M. Lee Allison, June 19, 2007, ESRI Users Conference, ArcGIS Geology Data Model town hall meeting, San Diego, CA

M. Lee Allison, November 2, 2007, OneGeology Management Meeting, Geological Survey of Canada, Ottawa, Canada

Proposals resulting from the workshop

M. Lee Allison, co-investigator, "Building a National Geoinformatics System," National Science Foundation, subcontract from University of Kansas, $5,000 (funded)

M. Lee Allison, Principal Investigator, "*Arizona Sample and Data Inventory,"* U.S. Geological Survey, Award number 07HQGR0124, $5,000, July 15, 2007 – June 30, 2008

M. Lee Allison, Co-PI, "*Geoinformatics: A Framework for Integrating Drought, Global Change, and Water Management Data,"* National Science Foundation, $19,650, (PI: Jacobs, UA, $250K), [not funded]

M. Lee Allison, Senior Personnel, "*Geoinformatics: Toward Developing a National Geoinformatics System*," National Science Foundation, $8,000, (PI: Walker, UKansas, $93,531), funded (10-29-07 verbal announcement)

M. Lee Allison, PI, "*INTEROP – Geoscience Information Network,"* National Science Foundation, $749,246, submitted (with Linda Gundersen, Tamara L. Dickinson, Stephen M. Richard)

M. Lee Allison and Linda C. Gundersen, Co-PIs, "*Data Trust Alliance"* (*DataNet preproposal)*, submitted January 2008 (PI: Francine Berman, San Diego Supercomputer Center, $20M)

Appendix A: Workshop Participants

Lee Allison,* Arizona Geological Survey
Sky Bristol, USGS Central Region Geospatial Information Office
Rod Combellick, Alaska Geological Survey
Mike Crane, USGS EROS Data Center
Beverly DeJarnett, Texas Bureau of Economic Geology
Tammy Dickinson,* USGS Program Coordinator,
National Geological and Geophysical Data Preservation Program
Dave Ferderer, USGS Central Region Energy Team
Kevin Gallagher, USGS Chief Technology Officer
Linda Gundersen,* Chief Scientist, Geology
Bruce Johnson, USGS Data Integration, Mineral Resources Program
Kathleen Johnson, USGS Program Coordinator, Minerals Resources Program
Karl Muessig, New Jersey Geological Survey
Annette Olson, USGS National Biological Information Infrastructure
Jay Parrish, Pennsylvania Geological Survey
Steve Richard, Arizona Geological Survey
Jingle Ruppert, USGS Energy Resources Program
Dave Soller, USGS National Geologic Map Database
John Steinmetz,* Indiana Geological Survey
Ron Teissere, Washington Division of Geology and Earth Resources
Harvey Thorleifson, Minnesota Geological Survey
Patrick Tucci, USGS Water Resources Discipline
Jerry Weisenfluh, Kentucky Geological Survey

* organizers

Appendix B: Workshop Agenda

Role of State Geological Surveys and U.S. Geological Survey in a Geological Information System for the Nation

Agenda: February 21-22, 2007, Denver Federal Center, Building 810, Powell Room

<u>Wednesday, February 21, 2007</u> (Breakfast on own. Coffee will be available at workshop site).

8:00 AM Welcome, rationale, and expectations-Linda Gundersen and Lee Allison

- The past (how we got here) and the future (where we want to go).
- How this workshop fits into the other activities planned in 2007 – March NSF Community Workshop, May *Geoinformatics 2007*, GSA sessions and others.
- Format for workshop – Discussion, presentations, and breakout groups. It is a working meeting which will result in a short position paper with recommendations for moving forward.

8:15 AM How would a coordinated framework of accessing and integrating data between the state geological surveys and USGS fit into the developing earth science cyberinfrastructure?

8:15 AM	USGS perspective	**Linda Gundersen**
8:30 AM	State Geological Survey perspective	**Lee Allison**
8:45 AM	Discussion on "perspectives talks"	

Workshop materials include a brief paper on the above question from each presenter. Discussion should focus on the roles of the USGS, NSF and state surveys with respect to a geoinformatics system. What geoinformatics activities are currently going on in each organization? What is NSF's policy/philosophy on funding geoinformatics activities, database development? What partnerships and agreements are needed?

<u>Outcome</u>: General agreement from the group on creation of a coordinated national framework for accessing and integrating state and USGS data and the roles and responsibility of each organization. Such a framework would be publicly available, distributed, and integrated.

Scribe: John Steinmetz 8-10

10:00 AM Break

10:30 AM The components of a geoinformatics system include archives, data services, catalog services, user applications, help system, and governance mechanisms (see attached list.)

- What defines a state/federal system? What are the products of it? How would it be managed/operated?
- What are the state and federal roles of the larger Earth science system?
- Are there core components that we can agree have a state/federal role?
- What are the state and federal roles with respect to these components?

USGS Lead – Kevin Gallagher
AASG Lead – John Steinmetz
Scribe – Linda Gundersen 10:30-12:00

Noon — Lunch on your own. See list and maps. There is also a cafeteria in Building 56.

1:30- PM — What are the opportunities and challenges to an integrated or coordinated effort among the state geological surveys and USGS?
USGS Lead - Tamara Dickinson
AASG Lead – Harvey Thorleifson
Scribe – John Steinmetz

- What are the lessons learned from USGS and state surveys?
  - **Dave Soller** talk about lessons learned from NGMDB (a centralized database). (15 mins)
  - **Jerry Weisenfluh** talk about lessons learned NatCarb (a distributed database). (15 mins)
- Is it possible to show the viability of a coordinated state and federal system by building a prototype for a few data components? This is a discussion of low hanging fruit – are there some things that could be done easily, cheaply, quickly to show progress? Or high priorities? Can adoption of a small set of protocols by the geological surveys create a data network that would serve as a catalyst for geoinformatics? What would these protocols be?
- Outcome: List of lessons learned and potential opportunities (low hanging fruit) that could demonstrate the concept and success.

3:30 PM — Break

4:00 PM — Opportunity and low hanging fruit breakouts. These will be determined by the results of discussion. Each group select a scribe and presenter.

5:00 PM — Adjourn Dinner at Jose O'Shea's at 6PM

<u>Thursday February 22, 2007</u>

8:00 AM — Summary of low hanging fruit breakout sessions and discussion – 10 mins each

9:15 AM — Challenges to design, build, and maintain data access and interoperability.

Technical Breakout

- What are the operational models?
- What are the technical issues to be resolved?

USGS Lead - Dave Ferderer
AASG Lead – Steve Richard
Scribe – Dave Soller

Policy Breakout

- How do we pay for this?
- How should we proceed?
- How do we get Surveys to participate?
- What are the necessary interactions with the larger Earth science community?

USGS Lead - Kate Johnson
AASG Lead – Jay Parrish
Scribe – John Steinmetz

10:15	Break
10:30	Plenary discussion of breakout sessions
Noon	Lunch on your own. See list and maps. There is also a cafeteria in Building 56.
1:30 PM	Review of agreed upon opportunities, system components, policy issues. Next steps and action items **Scribe-Linda Gundersen**
3:00 PM	Adjourn – Writing Team to meet to create first draft of report.

Appendix C: Glossary and Definitions

CGI – Interoperability Working Group of the *C*ommission for the Management and Application of *G*eoscience *I*nformation - a commission of the International Union of Geological Sciences (IUGS).

Client – a user service or interface application to access online data or tools. It may be a website, downloadable software, or purchased software. A portal is one type of client.

Community of practice – the concept that we learn not only as individuals but as communities. By engaging in communities of practice we increase our capacity and innovation as well as leverage our support for areas of interest. A community of practice is not merely a community with a common interest. But are practitioners who share experiences and learn from each other. They develop a shared repertoire of resources: experiences, stories, tools, vocabularies, and ways of addressing recurring problems. This takes time and sustained interaction. Standards of practice and reference materials will grow out of this experience. But the critical benefits include: creating and sustaining knowledge, leveraging of resources, and rapid learning and innovation.

Cyberinfrastructure - is envisioned as "comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments." (Atkins et al., 2003: http://www.nsf.gov/od/oci/reports/toc.jsp)

Data Integration Blueprint - long-term plan to achieve a USGS-wide vision of data integration and interoperability. Comprehensive, incorporating data integration and scientific tool development efforts of all USGS disciplines and regions into a single framework with common practices and a seamless infrastructure serving as the foundation for data integration across major USGS datasets, project-specific data sets and partner and collaborator data sets.

FGDC – *Federal Geographic Data Committee*

Geoinformatics – distributed, integrated digital information system and working environment that enables discovery of information, utilization of existing data, and access to applications for research, decision-making, and education.

GEON – *GEOlogic Network* – an NSF-funded consortium to develop geoinformatics, based at the San Diego Supercomputer Center, with 10 regionally based geologic testbed projects across the country.

GeoSciML – *GeoScience Mark-up Language* – a geoscience specific application that supports interchange of geoscience information between different database formats and software environments, and in particular for use in geoscience web services. GeoSciML is based on Geography Markup Language (GML – ISO DIS 19136) for representation of features and geometry, and the Open Geospatial Consortium (OGC) Observations and Measurements standard for observational data GeoSciML allows applications to use globally distributed geoscience data and information. GeoSciML is not a database structure. GeoSciML defines a format for data interchange. Data providers can provide a GeoSciML interface onto their existing data base systems with no restructuring of their internal databases required. Geoscience-specific aspects of the schema are based on the NADM conceptual model for geoscience concepts (2), and borehole information from the eXploration and Mining Markup Language (XMML).

GML – *Geography Mark-up Language* –

Grid - “A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.” (Carl Kessleman and Ian Foster, 1998, *The Grid: Blueprint for a new computing infrastructure*)

ISO – *International Organization for Standardization*

Interoperability – used here to describe the access to globally distributed digital data in standardized formats (regardless of database structure, operating system, or server configuration) that can be used immediately in applications.

Metadata – information about the nature of data, such as owner, origin, date, formats, etc (“data about data

NatCarb – National Carbon Sequestration Database project – a project funded by the US Department of Energy that created limited interoperability among databases from 130+ organizations in 47 states, 3 provinces, and 2 tribes, as part of 7 regional partnerships and provided applications for displaying and analyzing the results.

NSDI – *National Spatial Data Infrastructure* - program by the FGDC for creation of system-wide standards to disseminate government data uniformly

OGC – *Open Geospatial Consortium* -

Open source – the software code is freely available to users. Non-proprietary.

Portal – one type of client (or user interface)

Protocol – software used to convert one set of data or information into another, generally more common, standard. Markup languages are one type of protocols.

Schematic – describes data structures. An example is a mark-up language such as XML or GeoSciML. Considered ‘middleware.’

Semantic – describes data content, classification, and vocabularies. Ontologies are controlled vocabulary resources.

SOA – *Service Oriented Architecture* – a system designed to access or deliver data and software applications through the use of web-based services, standardized protocols, and interchange formats. It allows use of globally-distributed data and information in real-time. This contrasts with central data warehouses and desktop software for processing and analysis.

Syntax – describes how data are translated, using markup languages or other tools.

WCS – *Web Catalogue Service*

Web service – a software application that is offered and used online, as opposed to be installed on a user’s computer.

WFS – *Web Feature Service* - a web service for delivery of feature data in XML, such as a description of geometry that specifies location, with attribute data specifying properties associated with the geometry. An example would be the shape and description of geologic units cropping out.

WMS – *Web Mapping Service* - a web service for sending pictures or images (e.g. tif, jpeg) to the user's viewer client.

Wrapper – slang for software ("protocol") designed to translate data standards so they can be integrated with other data sources ("interoperability").

XML – *eXtensible Mark-up Language* –

XMML - *eXploration and Mining Markup Language*

Z39.50 – a standard used by the Library of Congress for digital bibliographic data.

Appendix D: Geoscience Information Network Steering Committee members

Co-Chair: Linda Gundersen, Chief Scientist, Geology, USGS
lgundersen@usgs.gov, 703-648-6601
Co-Chair: Lee Allison, State Geologist and Director, Arizona Geological Survey
Lee.allison@azgs.az.gov, 520-770-3500

AASG members

John Steinmetz, State Geologist and Director, Indiana Geological Survey
jsteinm@indiana.edu, 812-855-5067

Harvey Thorleifson, State Geologist and Director Minnesota Geological Survey
thorleif@umn.edu, 612-627-4780

Jay Parrish, State Geologist and Director, Pennsylvania Geological Survey
jayparrish@state.pa.us, 717-702-2053

USGS members

Anne Frondorf, Chief Scientist for Information
anne_frondorf@usgs.gov, 703-648-4205

Gladys Cotter, Associate Biology Director for Biological Information
gladys_cotter@usgs.gov, 703-648-4090

Kathleen (Kate) Johnson, Program Coordinator, Mineral Resources, USGS
kjohnson@usgs.gov, 703-648-6110

Peter Lyttle, USGS Landslide Hazard Program Coordinator, USGS
plyttle@usgs.gov, 703-648-6943