

U.S. Geoscience Information Network: A Critical Path for Data Integration in the U.S. Earth Sciences



Kevin T. Gallagher, U.S. Geological Survey, Reston VA and M. Lee Allison, Arizona Geological Survey, Tucson, AZ



U.S. Geoscience Information Network: A Critical Path for Data Integration in the U.S. Earth Sciences

Abstract:

Development efforts for the U.S. Geoscience Information Network (USGIN) over the last two years have crystallized around the Community for Data Integration (CDI) at the USGS, and the 50-state AASG State Geothermal Data project. The next step in developing a USGS-AASG community is to bring these two efforts into closer alignment through greater participation in CDI activities by geoinformatics practitioners from state geological surveys, and implementation of test bed activities by the USGIN partners, within two years.

Test bed activities in the geological survey community will define a scope and provide a foundation to promote the use of specifications developed by the larger geoinformatics community. Adoption of some of these specifications as 'standards' by USGS and AASG for use by those organizations will lend authority and motivate wider adoption. The arc from use case to test bed to production deployments to agreement on 'standard' specifications for data discovery and access must be propelled by active interest from the user communities who have a stake in the outcome. The specifications developed will benefit the organizations involved in development, testing and deployment, which motivates participation -- a model that has worked successfully for standards organizations such as OGC, ISO and OASIS.

The governance structure to support such a community process should promote grass root nucleation of interest groups that are the core of development efforts. Some mechanism for community agreement on priorities is desirable because geological survey agencies will need to allocate resources to support development. Loosely knit organizations such as ESIP and the current CDI provide models for this kind of structure. Because many geological surveys have data archive and dissemination functions as part of their portfolio, some support for the system can be built into the operating expenses and overhead. Sharing of resources and reuse of components can reduce the cost. Wide adoption of similar software, protocols and practices increases the number of stake holders with an interest in supporting the system.

USGIN – the roadmap



Evolution

- Conceived in 2007
- Meeting of representatives from
 - Association of American State Geologists (AASG)
 - US Geological Survey (USGS)
- Outcome
 - Develop a distributed national infrastructure of digital Earth Science data
 - Use common standards and protocols
 - Preserve data ownership, credit, and control
 - Build on existing systems

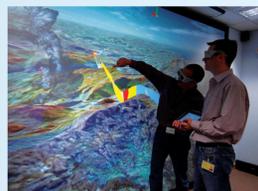


Purpose

- Use standardized services to make State and Federal geological survey data accessible online
- Employ a distributed network
- Use common standards and protocols
- Work with data providers to implement services

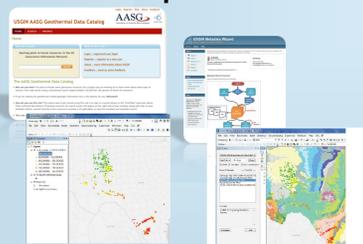
The Challenge of Data

1,000s of databases
1,000s of collections
80,000+ geologic maps



Converting this..... to this

Accomplishments



USGIN Catalog –

- Collection of consistent metadata describing network resources

USGIN Repository –

- Generates and manages metadata for files and Web tools
- Upload capability

Interoperable, Simple Content Models -

Widely utilized client platforms are available for service utilization (ArcGIS, Microsoft Oce, OpenLayers and other free, open-source software)

Web Services Deployment -

OGC Web Feature Services deployed by the Arizona, Illinois, Kentucky, Nevada geological surveys as server hubs for the NGDS.

Partnership Test Bed Activities

Test Bed Proposal 1: OneGeology – USGS

- USGS and AASG collaborate to implement Web Map and Web Feature Services according to the OneGeology profile
 - USGS: responsible for national maps (surficial deposits, glacial deposits, bedrock geology of US, etc.)
 - State Geological Surveys: deploy services for state-scale geologic maps
 - Community: agrees on an integrated portrayal scheme to attain interoperability between map services

- States and USGS deploy observation services for geochronologic data, geochemical data, gravity stations, or similar commonly available site or sample-based data
- Community agrees on
 - Service profiles and interchange formats
 - Procedures to avoid data duplication
- This Test Bed project could build on and enhance the National Geothermal Data System and EarthChem

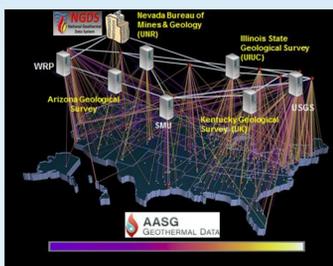
Test Bed Proposal 3: Integrated Catalog Capability

- Achieve interoperable metadata
- Demonstrate ability to harvest between catalog nodes hosted by state surveys and USGS
- Execute search against multiple nodes
- Edit tools to include user authentication and access control

Parallel tracks to building a national federated data integration framework



Design, build, deploy, and populate contributions to the National Geothermal Data System using USGIN framework



Digital data from 50 states relevant to geothermal energy exploration and development. >34,000 data sets so far.



USGS Community for Data Integration

CDI is a community of practice established in 2009

- Facilitate discovery of data and tools
- Improve use of scientific computing capacity
- Develop, implement scientific data products and services
- Enable data integration

CDI members

- USGS data providers/practitioners/scientists /consumers
- External partners in government, academia, industry

USGS Community for Data Integration

- CDI Working Groups
 - Technical Stack Working Group
 - Sub-teams:
 - Mobile Applications Development
 - Data Management Working Group
 - Sub-teams:
 - Data Policy
 - Data Best Practices



High-value opportunities for USGS:

- An application for better access to USGS corporate datasets through ArcGIS
 - Minerals Resources Data (MRDATA)
 - The National Map (TNM)
 - National Water Information Service (NWIS)
- A framework for loosely coupled models
- A tool for data upload, registry, and access
- A data management best practices Web site (in progress)
- A series of data management education modules (in progress)
- A new metadata policy to be published in the USGS "Survey Manual" (in progress)

A Partnership in Delivering Geosciences Information

USGIN Strategic Planning Working Group

- Develop a vision for USGIN
- Establish a roadmap for sustainability
- Define interoperability experiments among USGIN partners to demonstrate capabilities

Implement a Business Model:

- Establish USGIN as a non-profit foundation that will provide a conduit for financial resources
- Resource from user community and agency infrastructure support

Establish Governance:

- Community of Practice
- Technical Steering Committee (7 members)
 - Representation from State Geological Surveys, USGS, and other geoscience-focused communities
- 2-3 permanent staff



- Held several conference calls over 8 week period
- Gathered input from ESIP Federation Annual Meeting in a "Town Hall" format
- Developed a matrix of organizations with similar functions to compare business models, capabilities, governance, etc.
- Developed a Draft White Paper, currently in review

Outreach and Sustainability Planning:

- Engage State Geological Surveys
- Develop services and knowledge base that provides value to user community
- Design and plan activities so they have value to participants—learning, direct project application, sense of professional identity

Conduct 'Test Bed' Activities

- OneGeology – USGS
- Observation Services
- Integrated Catalog Capability

This material is based upon work supported by the Department of Energy under Award Numbers DE-EE0002850 and DE-EE0001120 and by National Science Foundation grant EAR-0753154



Thanks to Viv Hutchison & Cheryl Morris, USGS; Steve Richard, & Kim Patten, AZGS