# U.S. GEOSCIENCE INFORMATION NETWORK: DATA INTEGRATION FRAMEWORK

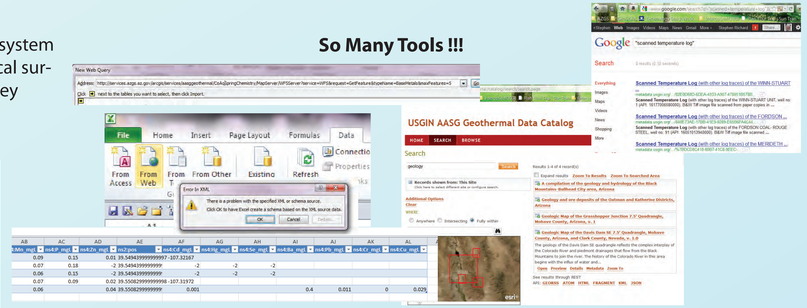
Ryan J. Clark and Stephen M. Richard
Arizona Geological Survey and USGIN steve.richard@azgs.az.gov

## What is the Network:

- -A collection of service definitions, interchange formats, and vocabularies
- -Catalogs that lead users to resources
- -A community of users: providers and consumers of data, services, applications

A national, distributed, interoperable system
Collaboration between state geological surveys and the U.S. Geological Survey

## So Many Tools !!!




## Progress in the past year

### Catalog

http://catalog.usgin.org/search/
ESRI Geoportal, free, open-source software (http://catalog.usgin.org/geoportal).
USGIN ISO metadata profile
OGC Catalog Service for the Web (CSW 2.0.2) and ESRI REST service interfaces.

### Document Repository

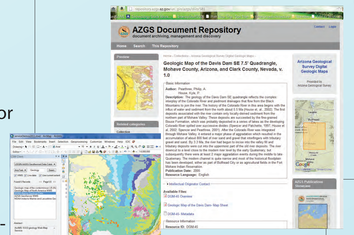
http://repository.usgin.org/
Web tool to create metadata and upload files to repository
Drupal v6 application (http://repository.usgin.org/);
Register and publish resources, pushes metadata to a web accessible directory for harvest into the catalog.

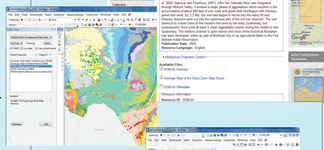
### Interoperable, Simple Content Models

http://www.stategeothermaldata.org/data_delivery/content_model_templates
Widely utilized client platforms are available for service utilization (ArcGIS, Microsoft Office, OpenLayers and other free, open-source software)
All State Geological Survey data contributed to the National Geothermal Data System uses these content models
Microsoft Excel spreadsheets to facilitate data loading;
Deployed services use an XML encoding of the content model as an interchange format.

### Getting Data Online

http://services.azgs.az.gov/ArcGIS/rest/services/
http://geothermal.isgs.illinois.edu/ArcGIS/rest/services/
http://kgs.uky.edu/ArcGIS/rest/services/
http://gisweb.unr.edu/ArcGIS/rest/services/
OGC Web Feature Services deployed by the Arizona, Illinois, Kentucky, Nevada geological surveys as server hubs for the NGDS.
AASG-USGS development of a relational database format for geologic map data (NCGMP09, http://ngmdb.usgs.gov/Info/standards/NCGMP09/).



## Promoting a sea change: Open data and flexible applications

tightly coupled client-server apps
roll your own, quick and dirty

gets done quick
fewer limitations
easier to start writing code
short life span
hard to maintain
duplication of efforts

loosely coupled client-server apps
documented, standards based

interoperable
easier to document
portable
reusable
don't maintain alone
takes longer
have to figure out someone else's code

## What works:

data online: simple feature services
search via CSW or Google
light weight tools for URI redirection
document repository (Drupal website)
Communities are developing

## What we're working on

metadata 'standards' aren't standard
getting feedback
documentation
standards development
quality assurance

## Bottlenecks

good metadata edit/maintenance workflow tools
training of workforce

## Service Inventory: Different formats for different uses

**Simple features**: browse, filtering, easy access. Processes that look at lots of generalized data — Atom, csw:record for metadata standard XML for data services, standard protrayal of map services

**Complex features**: rich, scientifically accurate content models. Processes that require smaller quantities of specific detailed information — GeoSciML, WaterML, SOS schemes

**Grid based representations**: single or small number of variables, large data extents, highly automated processing — NetCDF, WCS

## Machine actionable links:

**Does a computer know what to do with your URL?**
**A ubiquitous problem for next generation web**

| property | scope notes |
|---|---|
| **linkage** | URI identifies the resource that is the target of the link. (syn: href, targetURI) |
| **rel** | Semantics of link. (might need multiple values...) |
| **title** | Free text to label link in GUI |
| **type** | File format and optionally the native software application environment (MIME). |
| **function** | Term tells client why they'd use this link. e.g. 'download', 'browsing', 'fileAccess', 'service'. |
| **protocol** | URI protocol e.g. http, ftp, dns, nfs, See IETF registry at http://www.rfc-editor.org/rfcxx00.html. typically encoded as URI prefix |
| **serviceType** | URI that identifies a service protocol e.g. CSW, WFS, WMS, OpenSearch, OpenDAP, etc. |
| **outputScheme** | Identifier for schema of message retrieved by link; URI for xml schema or JSON scheme; Clients look at this to pick the link that will get a representation they can use; |

Other properties that may be useful (all optional)

| | |
|---|---|
| **altTitle** | String that encodes title value in a different character set, and/or contain language information as per [RFC5987]. |
| **descriptionurl** | Detailed text description of what the online resource is/does. |
| **media** | Indicates intended destination medium or media for style information e.g. 'screen', 'tty', 'print', 'braille,' 'aural'... Default to 'screen', and it is anticipated that other values would be only rarely required. |
| **length** | Indicates an advisory length of the linked content in octets; it is a hint about the content length of the representation returned |
| **hreflang** | Describes the language of the resource pointed to by the link |
| **behavior** | A comma separated list of properties specifying behavior expected in client when link is actuated or loaded; e.g. nofollow, showreplace, showembed, actuateOnLoad, actuateOnRequest, noreferrer, prefetch |

http://lab.usgin.org http://stategeothermaldata.org

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