

[Implementing USGIN, a Distributed Data Network for Geoscience Information](#) [1]

Submitted by superadmin on Fri, 2012-06-29 15:01 Tuesday, July 17, 2012 - 10:30 to 12:00

Event: [Summer Meeting 2012](#) [2]

Session Type: [Workshop](#) [3]

Media/Video: [session recording \(streaming\)](#) [4]

[session recording \(download\)](#) [5]

Expertise Level: [Intermediate](#) [6]

Identifier: doi:10.7269/P37P8W9Z

Collaboration Area: [Discovery](#) [7]

[Geospatial](#) [8]

[Information Technology and Interoperability](#) [9]

Abstract/Agenda:

The U.S. Geoscience Information Network (USGIN) is a framework for sharing geoscience data in a distributed network based on a collection of open-source applications, standards, procedures, and protocols for data sharing. USGIN is web-based, distributed, open-source, and interoperable. This workshop will provide a case study of USGIN in action through the Association of American State Geologists (AASG) State Geological Survey Contributions to the National Geothermal Data System.

Notes:

US GIN session

Demonstrating the USGIN through the contributions of the AASG. US Geoscience information network (USGIN) is for interoperability catalog services based on ISO and cataloging services standards.

Overview – what defines network, and introduction on USGIN, how we discover and serve data, and a live demonstration on USGIN.

What is a network?

Technology, resources and people that make up a network. With respect to data owners that still allow the community to grow. USGIN is a national network of digital data. It is collaboration between AASG (State geological surveys) and the USGS.

It is a network of data providers (the collaborators), data consumers and developers (hopefully commercial developers in the future). It has a few major hubs, including the Illinois State Geological Survey and Kentucky geological survey who have representatives here.

The focus is on selected data resources like well data – well logs, maps, sample data etc, to allow people more time to do science instead of maintaining/manipulating their data sets. They have a goal of putting the responsibility on the data provider in making data interoperable.

There are three tiers to access, we would like to get to the third:

First tier – putting data on line so that it can be discoverable. There is no interoperability but you can find it, search for it in a cataloging system.

Second tier – your data is discoverable (metadata is key) and providing interoperability – can ingest other's data in to your software system, but it is still in the original providers format.

Third tier – can analyze data across data sets. Can ingest because built on open protocols and can do analysis between different data sets.

“Who determines the data set content” and “Data integration” are two axis on this situation.

How hard is it, what are the obstacles? First getting the data (?); second tier – having data in the correct format, IT restrictions, setting up OGC service; third tier –developing an interchangeable format that everyone can agree on.

Stategeothermaldata.org – Consumers can discover, access and explore the resources. Also Data providers can share their resources.

In the future would like to have people register resources – create metadata and get it in a place where it can be accessed. They are also focusing on how to get people involved. Getting people like ESIP members to join meetings, there are many venues for this discussion. Also important is deploying a service – publishing data like maps etc. And are looking for people to write applications that can access USGIN catalog, like ARCGIS plugins and also working on an excel plugin as well. This is discussed on the lab development website.

They are using two metadata implementations – their own metadata recommendations and ISO 19139 and an xml transformation. An example was provided. There are three ways to create metadata, one – do it yourself. Two - they have an excel sheet with a python script that allows for easy upload of metadata with the appropriate xml, good for non-programmers. Three – using the online system.

Services for maps – there is a OGC webmap service which does not allow cross attribute search (that is what the web features are for).

Looking at creating a grid system for data sharing. And looking at a few different examples. These structured data formats - you have the observations (an event with data) and features (compiled and interpreted data).

The current interchange formats and are working on are listed on the slide, but are too long to type here – see slides for details. They have worked with the experts in these areas to find out what standards they are using.

USGIN – GML is trying to be kept simple. You can currently search a list of data in their beta version online.

USGIN is also using URIs for the data sets. This is important for long term access.

Questions?

Jennifer Davis asked about the different methods of metadata creation, which is more popular? Answer - They said that most of them are through excel spread sheets and transferred in to the catalog. They are working on making the metadata wizard (online system) more accessible.

Jennifer also asked about the validity check on the data – Answer – it is an automated process but yes, all data is checked.

Sarah asked about if the metadata is checked for validity as in completeness but also for syntax. Answer it is also visually checked, but they are checking for both at this time.

Ted asked about repeating fields in a spreadsheet? Answer – they delineate with a pipe symbol and Ted asked how it is then parsed when there is multiple entries for a resource – Answer, they will have to check.

Demonstration –

Discussion of the data sets handled by the system. See slides for list. They use XML for interchangeable formats. Next is transforming the excel spreadsheets submitted to the USGIN ISO XML. This is done with a python script on each spreadsheet added.

Metadata formats - ISO 19115, 19139, 19119, FGDC, OGC ETC

How do you contribute or become part of USGIN? Must have relevant data, must be in the required format, must have associated data and must have a way to share it.

They are currently developing a repository interface (it is in beta testing). It was briefly demonstrated to the group. But it has not yet been released to the public. They walked through some of the required fields, next they demonstrated a search for data in the collection.

They also mentioned they are working on naming the web space and if anyone has suggestions please pass them along.

Open to questions:

Matt asked about links to the CSW end point. Answer – they will get back to them on that.

Someone asked what the new catalog is based on. Answer, Jango and CouchDB. He followed up with questions about the services but the presenter was not sure at this time.

Matt asked about granularity with aggregated data sets or individual. Answer – depends on the record type. For example for the well logs – they have changed their mechanism for how it is provided. It is just metadata, and now they will be published as services, available by individual well log records. You can search them record by record.

Question – does that mean you have millions of records in your catalog? Answer – that includes in the catalog and under review. Clarification – metadata for every point? Yes.

[DEMO - How to deploy a webservice - \[S1\] \[S1\] \[10\]](#) this is done via a video

Using ARCGIS 10. They first opened a spreadsheet with data (with field headers displayed).

They opened arc catalog and navigate to where data is saved on the computer. Then they created a personal geodatabase (this allows you more manipulation when accessing it in [Access then excel. you can more easily validate the data types as well\[S2\] \[S2\] \[11\]](#)). Checking the feature classes to make sure everything was imported properly. They display the data in arcmap (it should auto populate the lat/long or you can select it from the drop down. WGS84 is usually used. Data was then exported, now as a map and web feature services (saved in the same database). As an aside, there are naming conventions for how these files are created/saved and they have been documented. The data is now added to the map, and the attribute table is demonstrated. This is followed by editing the layer properties to include additional metadata which can be harvested in to the catalog. Next is editing the map properties. The author is always credited here. Usually when they populate these, they pull the information from the original spreadsheet where the data was collected from. Just transferring it over. The project is then saved. It is then published to the ARCGIS server – and it is now up and running on the site. She also demonstrated how you can edit the data after you published it. The service is stopped, the properties are opened. Most of the details are in the capabilities document/section of the properties.

Another demo was given with the XML data files and how a data set is retrieved. *This was not done with a video, but specific notes were not taken.* An online ARCGIS interface was used to pull data from a specific server, from the URL in the XML file. Alternative processes for reviewing data were also used, specifically the USGS national map viewer. In the catalog system, you can also export an excel file of the attributes from the map selected so that you can have the data.

Matt asked a question –how do we define symbology for the WFS? Answer – we have a generic set so that they remain interoperable. We choose one and ask that people use the same one. Clarification, is this the one from the ESRI client? Answer – Yes. Someone commented that the client has to demonstrate the symbology.

Notes by: Sarah Ramdeen

Actions:

Do multiple entries get parsed separately in the metadata (from Ted Habermann)

Is there a link to CSW end point? Did they fill out service metadata for the csw end point (from Matt)

Session Leads:

Name: [Steve Richard](#) [12]

Organization(s): [Arizona Geologic Survey](#) [13]

Notes takers:

Name: [Sarah Ramdeen](#) [14]

Organization(s): [School of Information and Library Science UNC-CH](#) [15]

Email: ramdeen@email.unc.edu [16]

Name: [Kelly Monteleone](#) [17]

Organization(s): [University of New Mexico](#) [18]

Email: krbm@unm.edu [19]

Creative Common License: Creative Commons Attribution 3.0 License

Teaser: This workshop will provide a case study of USGIN, a framework for sharing geoscience data in a distributed network

Source URL: <https://commons.esipfed.org/node/447>

Links

[1] <https://commons.esipfed.org/node/447>

[2] <https://commons.esipfed.org/event/summer-meeting-2012>

[3] <https://commons.esipfed.org/session-type/workshop>

[4] <https://esipfed.webex.com/esipfed/ldr.php>

[5] <https://esipfed.webex.com/esipfed/lsr.php>

[6] <https://commons.esipfed.org/taxonomy/term/261>

[7] <https://commons.esipfed.org/collaboration-area/discovery>

[8] <https://commons.esipfed.org/collaboration-area/geospatial>

[9] <https://commons.esipfed.org/collaboration-area/information-technology-and-interoperability>

[10] https://00000129/#_msocom_1

[11] https://00000129/#_msocom_2

[12] <https://commons.esipfed.org/node/446>

[13] <https://commons.esipfed.org/taxonomy/term/312>

[14] <https://commons.esipfed.org/node/557>

[15] <https://commons.esipfed.org/taxonomy/term/373>

[16] <mailto:ramdeen@email.unc.edu>

[17] <https://commons.esipfed.org/node/430>

[18] <https://commons.esipfed.org/taxonomy/term/297>

[19] <mailto:krbm@unm.edu>