Metadata and the US Geoscience Information Network

Profiles, Software and Implementation

Background

USGIN, Interoperability and the National Geothermal Data System

Arizona Geological Survey 7/12/2011

The US Geoscience Information Network

- Partnership between the
  Association of American State Geologists (AASG)
  and the US Geological Survey (USGS)

- Objective is to make geoscience information easier to find, distribute, and analyze

- Build a nation-wide network for geoscience information that is:
  - Web-based
  - Open-Source
  - Distributed
  - Interoperable

The US Geoscience Information Network

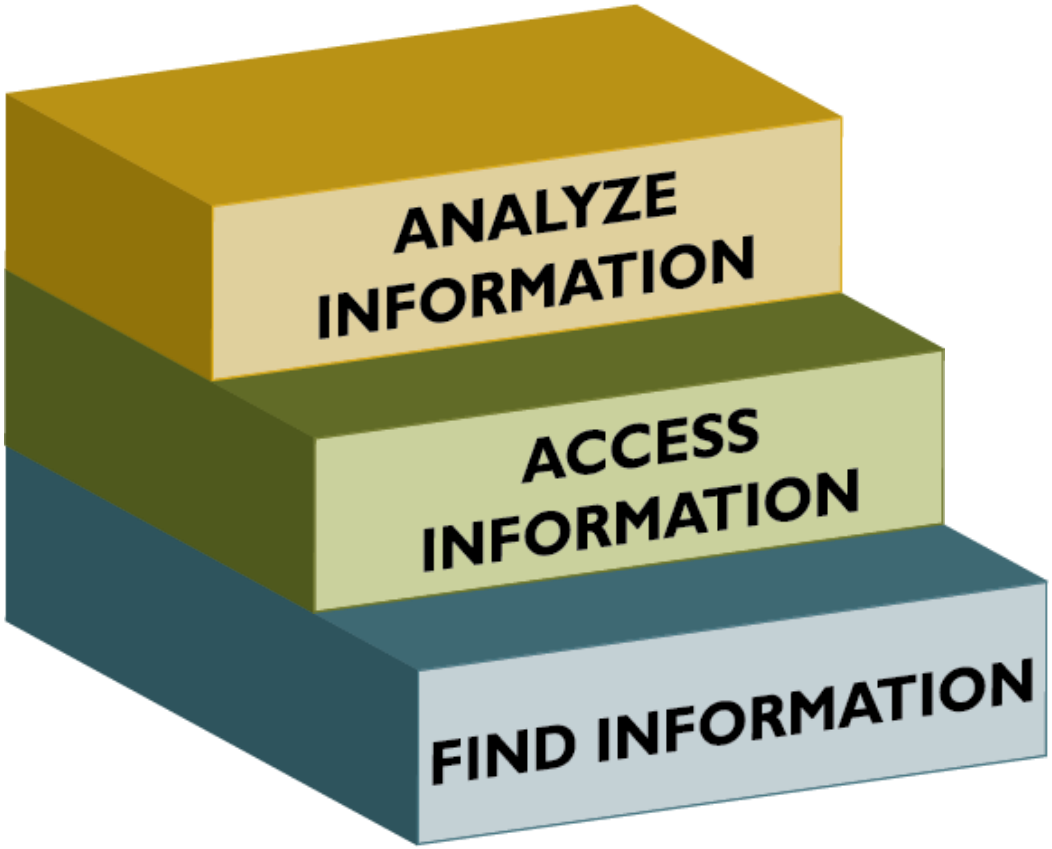
Distributed Network

- Data Clearinghouses are becoming obsolete
- Information is increasingly brought to us from disparate sources
- Constantly improving search capabilities allow us to find all this information
- Benefits of a distributed network:
  - Keeps information in the hands of the data providers
  - Allows for simpler update routines
  - Allows new information to be more rapidly conveyed to users

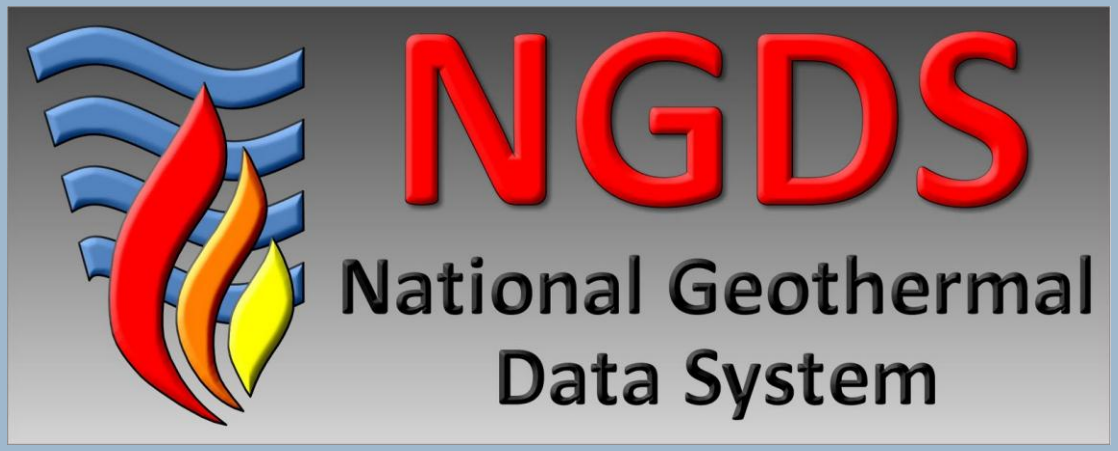
The US Geoscience Information Network

Interoperability… in Three "Easy" Steps

- Clearinghouse-style "portal" mapping applications are dead …
  instead we want mash-ups that work
- Analyze: Provide data services in dataset-specific standardized schema
- Access: Provide resources themselves online using standard OGC protocols
- Find: Provide standardized metadata for resources that may or may not be available online



The National Geothermal Data System



DOE & USGS Data

Boise State University

National Assessment

USGS

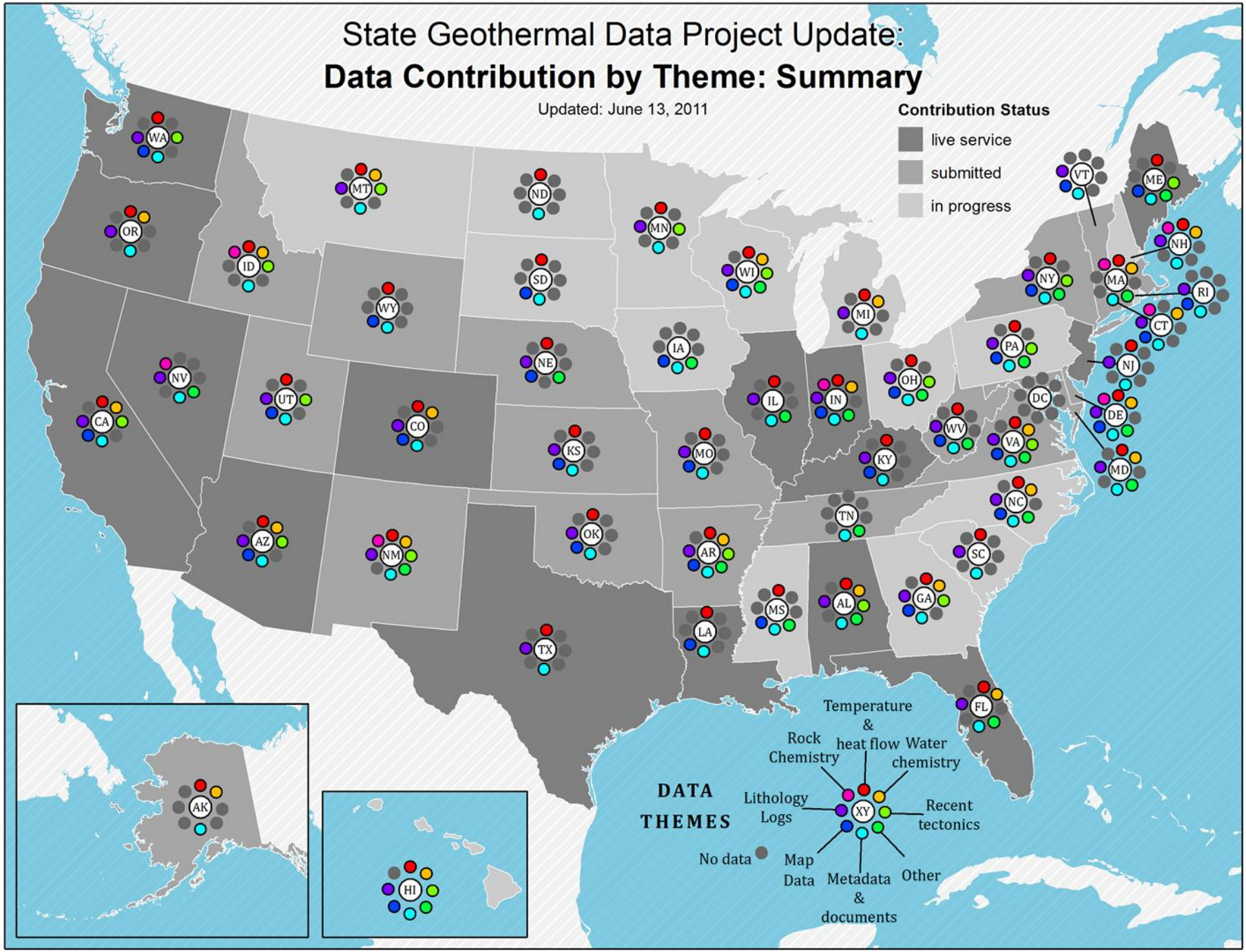
University Data

Southern Methodist University

State Geological Survey Contributions to the NGDS

AASG - AZGS

DOE Geothermal Technologies Program-funded projects



Project sponsored by the U.S. Department of Energy under award DE-EE0002850 to the Arizona Geological Survey acting on behalf of the Association of American State Geologists. Cartography by JCD.

Metadata Implementation

USGIN Profile and Custom Software

USGIN Metadata Implementation

Why Google Is Not Enough

- Search Efficiency
  - Good with text, not so good with data
  - Steps in the right direction: KML and GeoRSS indexing
  - Another approach: provide textual descriptions of data sets that can be indexed. For example, ArcGIS Server REST endpoints
  - ArcGIS Online, GeoNode ???
- Query Complexity
  - Easier to answer complex questions with structured metadata

Find geo
scale <
Iron M

Find bo
penetrate
formation

Find locations for samples with uranium-lead geochronology data in a given area.

… For now, we're stuck with formal metadata …

USGIN Metadata Implementation

- Minimum Metadata Recommendations

Title (1 entry): Succinct (preferably <250 characters) name of the resource.
Description (1 entry): Inform the reader about the resource's content as well as its context.
Originators (1 to many entries): Authors, editors, or corporate authors/curators of the resource.
Publication Date (1 entry): Publication, origination, or update date (not temporal extent) for the resource. Use a "year" or ISO 8601 date and time format. Alternative date formatting must be machine readable and consistent across all datasets. If no publication date is known, estimate the publication date range, enter the oldest year as the publication date, and include the estimated date range in the Description field.
Geographic Extent - *Horizontal* (1 entry, point or minimum bounding rectangle): Values given in decimal degrees using the WGS 84 datum. Required if resource has location Some resources may not be usefully described by an extent; if no extent is specified the default is Earth.
Access Instructions (1 entry): Text description of how to access the resource.
Distribution Contact Party (1 entry): The party (name of organization or person, etc.) to contact about accessing the resource.
Distribution Contact Email (1 entry): How to contact the party responsible for distribution
Metadata Date (1 entry): Last metadata update/creation date-time stamp in ISO 8601 date and time format. This may be automatically updated on metadata import if a metadata format conversion is necessary.
Metadata Contact Party (1 entry): The party (name of organization or person, etc.) to contact with questions about the metadata itself
Metadata Contact Email (1 entry): How to contact the party responsible for metadata content and accuracy
Metadata Specification (1 entry): Identifier for metadata specification used to create a metadata record encoding this content.

ments for USGIN metadata. The intention is to reduce the daunting complexity of the ISO

- USGIN Profile for ISO 19139 Metadata

```
▼<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
 xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:gml="http://www.opengis.net/gml"
 xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instan
 xsi:schemaLocation="http://www.isotc211.org/2005/gmd
 http://schemas.opengis.net/csw/2.0.2/profiles/apiso/1.0.0/apiso.xsd">
 ▼<gmd:fileIdentifier>
    <gco:CharacterString>3c529816-a0f4-11e0-8768-0024e880c1d2</gco:CharacterString>
  </gmd:fileIdentifier>
 ▼<gmd:language>
    <gco:CharacterString>eng</gco:CharacterString>
  </gmd:language>
 ▼<gmd:characterSet>
    <gmd:MD_CharacterSetCode
    codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resou
    codeListValue="utf8">utf8</gmd:MD_CharacterSetCode>
  </gmd:characterSet>
 ▼<gmd:hierarchyLevel>
    <gmd:MD_ScopeCode
    codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resou
    codeListValue="dataset">dataset</gmd:MD_ScopeCode>
  </gmd:hierarchyLevel>
 ▼<gmd:hierarchyLevelName>
    <gco:CharacterString>Dataset</gco:CharacterString>
  </gmd:hierarchyLevelName>
 ▼<!--
    Metadata point of contact - Point of contact for the metadata record, e.g. for users to
    report errors, updates to metadata, etc.
  -->
 ▼<gmd:contact>
  ▼<gmd:CI_ResponsibleParty>
   ▼<gmd:individualName>
      <gco:CharacterString>Missing</gco:CharacterString>
    </gmd:individualName>
   ▼<gmd:organisationName>
      <gco:CharacterString>Arizona Geological Survey</gco:CharacterString>
    </gmd:organisationName>
   ▼<gmd:contactInfo>
    ▼<gmd:CI_Contact>
     ▼<gmd:phone>
      ▼<gmd:CI_Telephone>
       ▼<gmd:voice>
          <gco:CharacterString>520-209-4119</gco:CharacterString>
        </gmd:voice>
      </gmd:CI_Telephone>
     </gmd:phone>
     ▼<gmd:address>
      ▼<gmd:CI_Address>
       ▼<gmd:deliveryPoint>
          <gco:CharacterString>416 W. Congress, Suite 100</gco:CharacterString>
```

USGIN Metadata Implementation

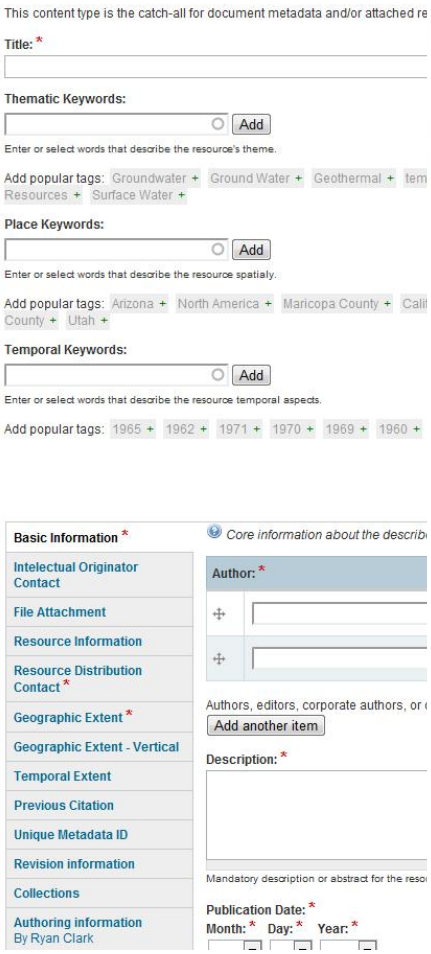
Metadata Creation and Inclusion in a Catalog

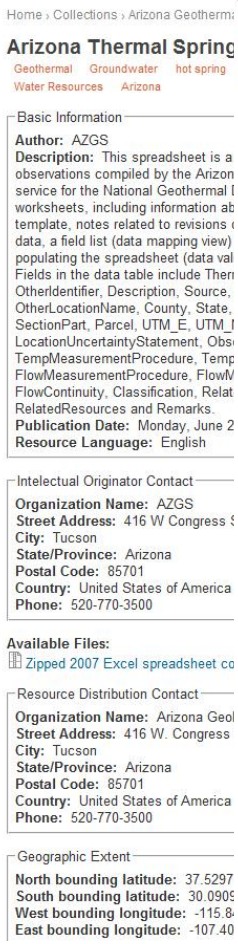
Four Options:

1. Write XML, upload to Catalog
2. Excel Sheet + Python
   - For bulk updates, ETL
3. Metadata Wizard
   - For offline resources or resources already online
4. Document Repository
   - For resources that need to be made available online

USGIN Metadata Implementation

Drupal-based Metadata Creation





Catalog Implementation

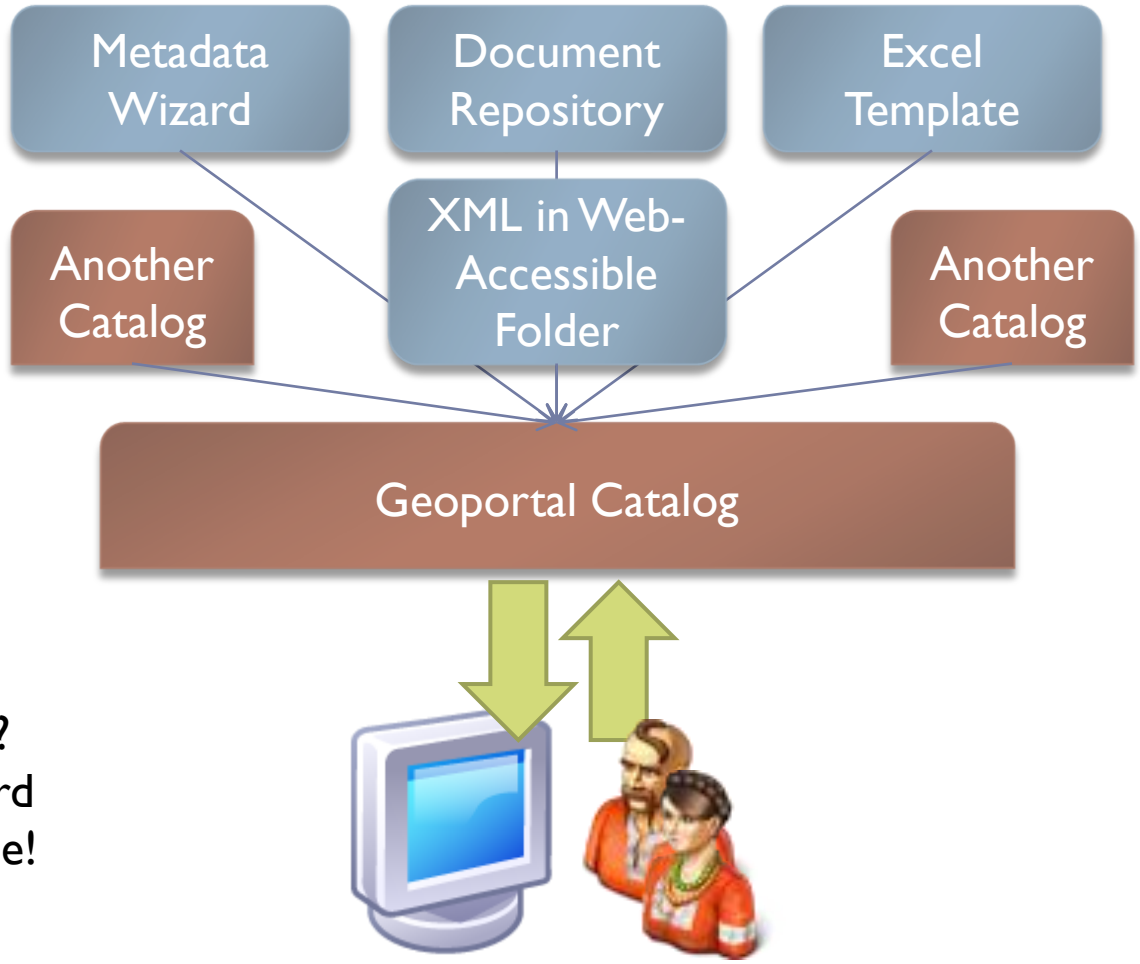
Using ESRI's Geoportal Server

Arizona Geological Survey 7/12/2011

USGIN Catalog Implementation

Why a Geoportal Catalog?

- Allow a variety of metadata creation methods to be aggregated
- Allow distributed metadata records to be aggregated: harvesting between catalogs
- Provide a consistent interface for searching and retrieving metadata records: CSW
- Why ESRI's Geoportal Server?
  - Geonetwork was too hard
  - They made it open source!



USGIN Catalog Implementation

Implementing a Custom Profile

- Serious about standardized metadata: removed profiles supporting any other format
- Wrote new definition and template files for our ISO profile, pretty much from the ground up
- New XSLT transformations allow harvesting WMS/WFS GetCapabilities directly into the catalog
- property-meanings, indexables, definitions:
  - mapping from data structure to indexed terms
  - potential semantic mediation tool???

USGIN AASG Geothermal Data Catalog

AASG Association of American State Geologists

USGIN

Home

You can simply...

Starting point to locate resources in the US Geoscience Information Network:

Search

What is this site?...

Login - registered user login

Register - register as a new user

About - more information about USGIN

Feedback - send us some feedback.

The AASG Geothermal Data Catalog

- **Why are you here?** You want to locate some geoscience resources for a project you are working on or learn more about some topic of interest. Your web search using a commercial search engine yielded 210,026 hits, 90 percent of which are nonsense.
- If you are looking for geothermal-related geologic information this is the website for you. **Welcome!!!**
- **How do you use this site?** The easiest way to get started using this site is to type in a search phrase in the "Find Data" input box above. Titles and brief descriptions of located resources for search results will appear on the right side of your window, along with links to view metadata details, connect directly to the resource or preview it (if applicable), or view the complete xml metadata record.
- **This site** uses the **ESRI Geoportal Server open source project (v. 10)**

What is this repository?

At the heart of the Geoscience Information network is a catalog system that enables data providers to publish metadata for data and services, and for data consumers to discover those resources. The catalog system is based on a federated system of metadata registries (databases that host and manage the metadata) that are accessible for search and harvest via public web services. Metadata records in this system are viewed as a public resource. For more information see **USGIN Catalog Service**.

This site is an access point to search the catalog. We are collecting metadata for geoscience resources, especially those geographically associated with the United States, and will be adding continuously to the metadata store. If you have resource metadata you would like to distribute to the community please **contact the USGIN project** or visit **the USGIN website.**

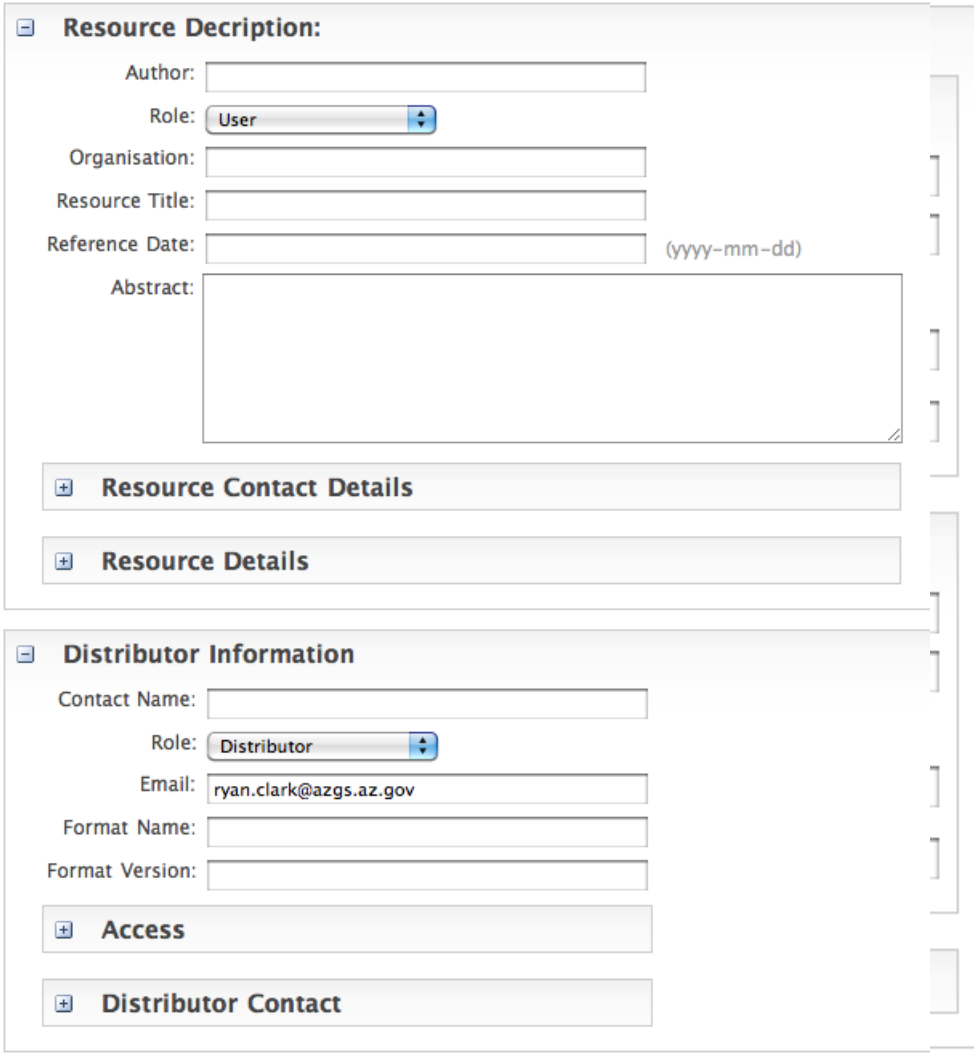
USGIN Catalog Implementation

Hardships

- Metadata creation user-interface limitations
- Validation: When schema-valid is not enough

http://gitorious.org/usgin/xmlvalidator/
http://gitorious.org/usgin/usginvalid/

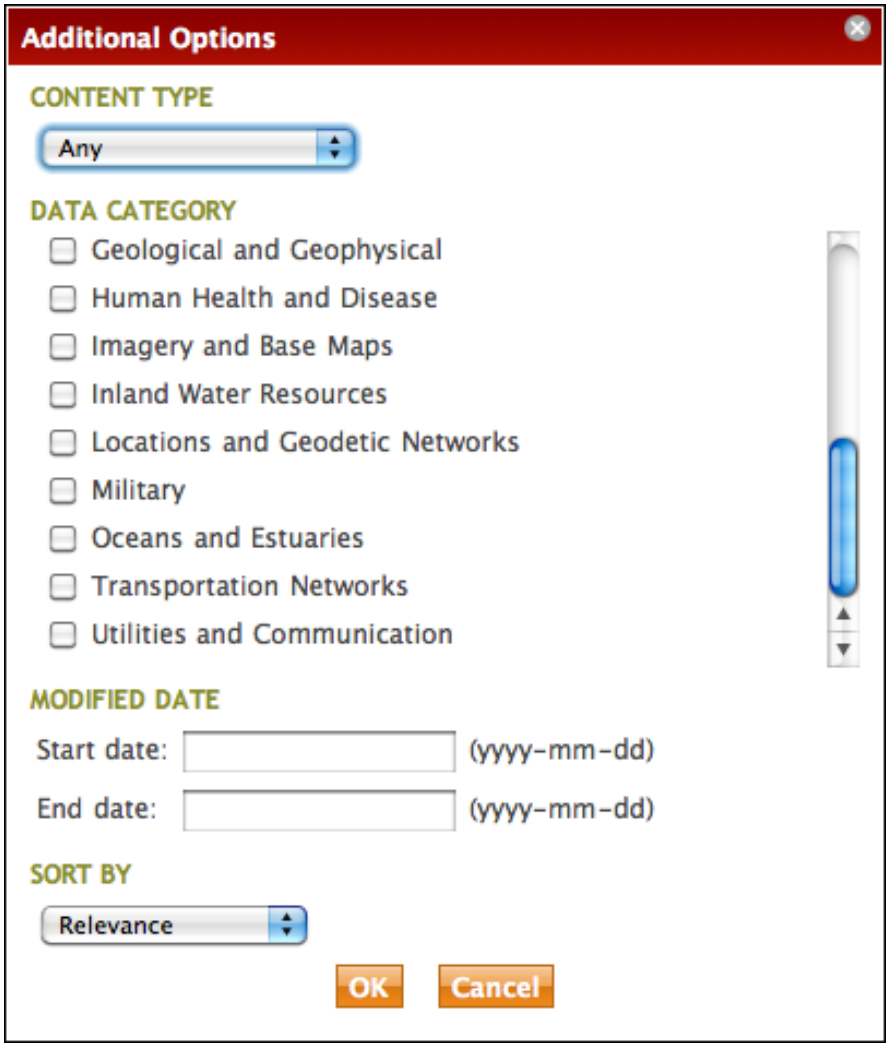
- FGDC to ISO XSLT requires XPath 2.0 support



USGIN Catalog Implementation

Future Direction

- What about complex queries?
- Upgrade to the new version
- Map-centric search: UI Design



Thank You

Ryan Clark
Arizona Geological Survey
ryan.clark@azgs.az.gov