USGIN
U.S. Geoscience Information Network

Metadata Recommendations for Geoscience Resources

Title:
Metadata Recommendations for Geosciences Resources

Latest released version:
pre-release

Creator:
USGIN Specifications Drafting Team and NGDS developer team

Editors:
Stephen M. Richard, Wolfgang Grunberg

Creation date:
2010-03-10

Last revision date:
2011-08-26

Document Status:
V. 1.04, presented for review and adoption

Publisher:
Arizona Geological Survey, Boise State University

Description:
This document provides guidance on the metadata content required to meet the use requirements for USGIN metadata. The intention is to reduce the daunting complexity of the ISO metadata specifications to a manageable level to promote development of interoperable metadata records for a federated resource catalog system.

Contributor:
See acknowledgements

Document Identifier:
USGIN2011-002

Notices

Neither the USGIN project, nor any of the participating agencies, take any position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither do they represent that there has been any effort to identify any such rights.

This document and the information contained herein is provided on an "AS IS" basis and USGIN DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Revision History

Version	Date	Comments	By
0.1	2010-03-10	Initial draft document	Stephen Richard
1.0	2010-07-15	Insert field summary developed for AASG Geothermal data production project, with edits. Identify minimum metadata record content and recommended content. Remove 'minimum' from title.	Stephen Richard
1.01	2010-07-19	Review, title change, and formatting	Wolfgang Grunberg
1.02	2010-07-28	Review	Stephen Richard
1.03	2010-07-28	Formatting, PDF preparation	Wolfgang Grunberg
1.04	2011-08-26	minor revisions to update text	Stephen Richard

Acknowledgement

Many individuals and organizations have contributed to or inspired the development of these Guidelines.

The USGIN Specifications Drafting Team include (alphabetically):

Ryan Clark – Arizona Geological Survey (AZGS)
Wolfgang Grunberg – AZGS
Stephen M. Richard – AZGS

The NGDS Development Team include (alphabetically):

Kim Kurz – Boise State University (BSU)
Christian Loepp -- BSU
Walter Snyder – BSU
Jordan Hastings – Nevada Bureau of Mines and Geology

Funding Provided by (chronologically):

National Science Foundation under EAR-0753154 to the Arizona Geological Survey acting on behalf of the Association of American State Geologists; 2009.
US Department of Energy under award DE-EE0001120 to Boise State University; 2010.

Contact Information

Arizona Geological Survey
416 W. Congress St., Suite 100.
Tucson, Arizona 85701-1381
Phone: 520.770.3500
Fax: 520.770.3505

Email

metadata@usgin.org

Online

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

Table of Contents

Tables and Figures

1 Introduction

A key component of a distributed information network is a catalog system that allow data and service providers to register resources, and data consumers to locate and use those resources. Such a system consists of a metadata register-- a collection of data items that describe available resources, a search interface that allows users to search the metadata register, and registration processes that enable resource providers to create or submit new metadata records to the register. Currently, many online catalogs are web pages with collections of URLs for services, or services are discovered accidently or by word of mouth. The vision is to enable a web client (portal) to search across one or more metadata registries without having to configure the client individually for each of the registries that will be searched. Thus, metadata providers can focus on data development, without having to also develop web clients to enable search of that metadata.

Production of quality metadata is time consuming, tedious, and gets little recognition, but good metadata is an essential component of a useful federated information system. Existing metadata standards are large complex information schema designed to account for any kind of resource description someone might want to create. This complexity makes them hard to use. Our goal is to define minimum content requirements that can be described in relatively simple language with common sense explanation of what the purpose of the content is. The scoping of the requirements is based on a collection of scenarios for how the metadata is intended to be used.

1.1 Normative References

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

USGIN_ISO_Metadata_1.1.3 USGIN profile of ISO content models (ISO19115 and ISO19119) and encoding (ISO19139). Access at http://lab.usgin.org/node/235.

ISO 19115 designates these two normative references:

- ISO 19115:2005, *Geographic information - Metadata*
- ISO 19115/Cor.1:2006, *Geographic information – Metadata, Technical Corrigendum*

ISO 19119 designates these normative references:

- ISO 19119:2005, *Geographic information - Services*
- ISO 19119:2005/Amd 1:2008, *Extensions of the service metadata model ISO 19108 designates*:
- ISO 19108:2005, *Geographic information – Temporal Schema*

ISO/TS 19139:2007, Geographic information - Metadata – XML Schema Implementation

ISO 10646-1, Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane

RFC 2119, Key words for use in RFCs to Indicate Requirement Levels, Network Working Group, 1997.

1.2 Purpose

This document is intended to provide guidance on the metadata content required to meet the use requirements for USGIN metadata. The intention is to reduce the daunting complexity of the ISO metadata specifications to a manageable level to promote development of interoperable metadata records for a federated resource catalog system.

1.3 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in Internet RFC 2119.

Application profile: a schema that consists of data elements drawn from one or more namespaces, combined together by implementers, and optimized for a particular local application. (Rachel Heery and Manjula Patel, 2000, http://www.ariadne.ac.uk/issue25/app-profiles/)

Catalog application: Software that implements a searchable metadata registry. The application must support the ability to register information resources, to search the registered metadata, to support the discovery and binding to registered information resources within an information community.

Codelist (also as Code list): a controlled vocabulary that is used to populate values for an xml element.

Data product specification: a definition of the data schema and value domains for a dataset. The data schema specifies entities (features), properties associated with each entity, the data type used to specify property values, cardinality for property values, and if applicable, other logical constraints that determine data validity. Value domains are specified for simple data types—strings or numbers, and may include controlled vocabularies for terminology required to specify some properties.

Dataset series: collection of datasets sharing the same product specification (ISO 19115). ISO 19115 does not define product specification. For the purposes of USGIN, a product specification defines a data schema, any required controlled vocabularies, and recommended practices for use of schema (see Data product specification).

Dataset: an identifiable collection of data (ISO19115). USGIN refines this concept to represent a collection of data items in which individual data items are identified and accessible. USGIN extends the concept of data items to include physical artifacts like books, printed maps and diagrams, photographs, and material samples--any identifiable resource of interest. DCMI definition is "Data encoded in a defined structure" with additional comment "Examples include lists, tables, and databases. A dataset may be useful for direct machine processing." Metadata for the collection is a different type than metadata for individual items in the collection (dataset vs. features). Criteria for what unifies the collection are variable (topic, area, author...). Data items may represent intellectual content -- information content and organization (data schema) -- or may represent particular manifestations (formats) of an intellectual artifact.

Interoperability: "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." ISO/IEC 2382-01 (SC36 Secretariat, 2003)

Metadata element: a discrete unit of metadata (ISO 19115), an attribute of a metadata entity. A metadata element contains some content specifying the value of the element; this content may be simple—a number or string, or may be another metadata entity.

Metadata entity: a named set of metadata elements describing some aspect of a resource.

Metadata register: an information store that contains a collection of registered metadata records, maintained by a metadata registry. (ISO 11179)

Metadata registry: an information system for assignment of unambiguous identifiers to administered metadata records. (ISO 11179)

Metadata section: Part of a metadata document consisting of a collection of related metadata entities and metadata elements (ISO 191115).

Metadata: data about a resource in some context. Generalize from ISO 11179 definition of metadata, which constrains the scope to data about data. For USGIN purposes, metadata may describe any resource—including electronic, intellectual, and physical artifacts. Metadata represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software.

Profile: set of one or more base standards and - where applicable - the identification of chosen clauses, classes, subsets, options and parameters of those base standards that are necessary for accomplishing a particular function [ISO 19101, ISO 19106]

Resource: An identifiable thing that fulfills a requirement. Usage here is closer to definition used in RDF (www.w3.org/TR/REC-rdf-syntax), generalized from ISO19115, which defines resource as an 'asset or means that fulfills a requirement'. Another definition is "An object or artifact that is described by a record in the information model of a catalogue" (OGC 07-006r1), but we broaden the intention to include any object or artifact that *can be* described by a record…

Service metadata: metadata describing the operations and information available from a server.

Source Specification: The specification or standard that is being profiled.

User Community: A group of users, e.g. within a supply-chain industry, the members of which decide to make a similar usage of the source specification in order to be able to interoperate.

Note that throughout this document, the names of xml elements are shown in `this typecase`. Long X-paths have been broken with non-breaking hyphen characters. Note that hyphens are not used in any xml attribute or element name, so if they appear in the text, they are strictly for better text wrapping. In Xpath expressions `/../` indicates that some elements have been omitted from the path.

2 Abbreviations

CSW	Metadata Catalog for the Web. Also abbreviated as CS-W and CS/W
GeoSciML	Geoscience Markup Language
GML	Geographic Markup Language
GUID	Global Unique Identifier
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
UML	Unified Modeling Language
URI	Universal Resource Identifier
USGIN	U.S. Geoscience Information Network
WCS	Web coverage Service
WFS	Web Feature Service
WSDL	Web Services Description Language
XML	eXtensible Markup Language

3 Use cases, scenarios, requirements

This section includes a number of user scenarios for how we intend USGIN metadata to be used, and discussion of several basic approach requirements that guide metadata content recommendations. At its heart, the problem is to find resources of interest via the internet, based on criteria of topic, place, or time, evaluate resources for an intended purpose, and learn how to access those resources. Detailed metadata describing a resource data schema, describing service or application operation, or providing detailed descriptions of analytical techniques and parameter are outside the scope intended for USGIN metadata. Our contention is that this more domain/resource specific type information is better accounted for with linked documents utilizing schema appropriate to those specific resource. Some examples include OGC getCapabilities, WSDL, and ISO 19110 feature catalogs.

- A user specifies a geographic bounding box or one or more text keywords to constrain the resources of interest, and searches a metadata catalog using these criteria. The user is presented with a web page containing a list of resources that meet the criteria, with links for each resource that provide additional detailed metadata, and direct access to the resource if an online version is accessible, e.g. as a web page, Adobe Acrobat document, or online application (see Accessing Resources, below).
- A client application provides user with a map window that contains some simple base map information (political boundaries, major roads and rivers). User wishes to assemble a variety of other data layers for a particular area for some analysis or data exploration, e.g. slope steepness, geologic units, bedding orientation, and vegetation type for a hazard assessment. User centers map view on area of interest, then using an 'add data' tab, accesses a catalog application that allows them to search for web services that provide the desired datasets. After obtaining the results and reviewing the metadata for the located services, user selects one or more to add to the table of contents for the client application. Response from catalog has sufficient information to enable the client application to load and use the resource (e.g. serviceType, OnlineResourceLinkage). More concrete instances of this case would be finding Web Map services to add as layers in an ESRI ArcMap project, borehole Web Feature Services to post borehole logs in a 3-D mapping application, or water chemistry data Web Feature Service to bring data into a spreadsheet or database.
- User searches for boreholes in an area. Returned metadata records have links to metadata for related resources, like logs of different types, core, water quality data, etc. that the user can follow to browse metadata for these resources.
- A catalog operator wishes to import and cache catalog records from a collaborating catalog that have been inserted or updated during the last month (harvest). This operation requires knowledge of the metadata standard and version used for the returned records.
- A user discovers an error in a metadata record for a resource that they have authored, and wishes to contact the metadata producer to request correction.
- A search returns several results that appear to contain the desired content, and user must select the most likely to meet their needs. Metadata should provide sufficient information to guide this decision.
- A project geologist at Company X is searching for data relevant to a new exploration target, and wishes to restrict the search to resources that are publicly available.
- Complex search examples (see further discussion in the Query complexity section, below):
    - Search based on related resources, for example a search for boreholes that have core.
    - Boreholes that penetrate the Escabrosa formation.
    - Sample locations for samples with uranium-lead geochronologic data.
    - Find links to pdfs of publications by Harold Drewes on southeast Arizona.
    - Find geologic maps at scale < 100,000 in the Iron Mountains.
    - Who has a physical copy of USGS I-427?

3.1 Efficient searching

A search should return results that are actually relevant. Existing web search tools are very good at indexing relevance based on association of words in text, and using links and user navigation history for those links. This kind of indexing does not work for datasets, in which the information may be encoded in binary format, and proximity of strings may be a function of the data serialization algorithm, not the semantics. Semantic technology is advancing rapidly, and there is significant effort devoted to increasing search efficiency using background information (common sense) encoded in ontologies. To index structured data more effectively and take advantage of semantic technology, users must describe resources using controlled vocabularies (ideally linked to an ontology) in a formal metadata schema. Practically speaking, semantic technology is still in its infancy (maybe early childhood?), but the issue is important for discovery of structure data. Thus, use of controlled vocabularies for metadata content that is meant to enable search for particular resource characteristics is a requirement. Determining the elements requiring such vocabularies must be based on specific use cases.

3.2 Identifiers

A widely used identifier scheme is important to reduce duplication, and determine associations between resources. Globally unique identifiers are essential for the described resource, and for the metadata record. The USGIN URI scheme (http://repository.usgin.org/uri_gin/usgin/dlio/331) has been put forward as a proposal for use in the US Geoscience Information Network.

The current thinking in the WWW community appears to be converging on a consensus to use HTTP URIs that are expected to dereference to some useful resource representation. A widely used and understood identifier scheme also enables semantic web functionality; “anyone can say anything about anything” requires being able to identify the things. Of primary interest here are crowd-sourced tagging of resources and feed back on utility, and related resources.

3.3 Query complexity

The complex search examples in the use cases section involve associations between resources, or resource-specific properties. The following table is a decomposition of some complex query examples

Table 1. Analysis of complex queries

Case#	Plain language query	Decomposition	Simplified solution
1	Boreholes that have core in a particular depth interval in a given area.	Borehole-centric approach -- geographic search for borehole resources (assume collar location), filter for those that have a related resource ‘core’, filter again for property of related resource ‘core interval = min, max depth meters’. Alternatively, view search as actually for a ‘core’ resource, so search should be for ‘core’ with some given vertical extent. The core resource must provide an ID ‘xxxx’ for the borehole from which it was obtained. To obtain more details about the borehole, search for metadata on borehole with resource ID = ‘xxxx’.	Include keywords for other resources associated with borehole. Put information about these in the abstract. User searches catalog for borehole with keyword (thesaurus=related resource) = ‘core’, reads abstract to see if it is what they want. The keywords would have to be a controlled vocabulary.

Case#	Plain language query	Decomposition	Simplified solution
2	Boreholes that penetrate the Escabrosa formation in a given area.	Geographic search for borehole resources (assume collar location), filter for property 'intersects Escabrosa formation'. Alternatively, search for borehole service that includes property = "formation tops", then query that service. Service properties would have to be from controlled vocabulary.	Include names of penetrated formations as keywords on a borehole. Formation names ideally from a geologic unit lexicon.
3	Locations for samples with uranium-lead geochronologic data in a given area.	Search catalog for Geochronology data service with property = 'analysis type' and backtrack to location point through sample metadata, or search catalog for U-Pb Geochronology Data Service and backtrack to location point through sample metadata, or search for 'sample service' with property = 'analysis type'. In the second case, there would still need to be some metadata property to indicate the analysis type for the service. Approach via the analytical data service requires chaining to the sample feature service, analogous to case 1 for borehole service.	Include keywords for kinds of analytical data associated with a sample in the sample metadata record. Search for samples with keyword (thesaurus=analysis type) = 'U-Pb geochronology'.
4	Find links to pdfs of publications by Harold Drewes on southeast Arizona.	Search for document resource with author = 'Harold Drewes' and geographic extent = 'SE Arizona', and online distribution format = 'pdf'.	Same
5	Find geologic maps at scale < 100,000 in the Iron Mountains.	Search for geologic map resource with geographic extent = 'Iron Mountains, and resolution scale denominator < 100000.	Same
6	Who has a physical copy of USGS I-427?	Search for document publisher = USGS, Series ID = I-427, offline distribution format = 'paper copy'	Include the document ID in the resource description.

Consideration of these queries indicates a requirement to distinguish metadata service from a data service. When the request involves properties of specific instances of a particular resource type, a data service for that resource should be accessed. The metadata for that service should describe the properties offered for resource instances in that service.

Cases 1-3 can be handled in a general way by a service chaining process, in which the catalog is searched for services offering the feature of interest with the property of interest that will be used as a selection criteria. This approach keeps the top level resource catalog simpler, but makes discovery operations significantly more complex. Cases 1-3 can also be handled with scoped keyword terms, where the scope includes things like 'analysis type', 'geologic unit', 'related resource type'. In this usage, the scope specifies a controlled vocabulary of categories related to some concept. Addition of new querying capabilities requires adding additional scoped keywords in the metadata. The second approach is viewed as more appropriate in a 'keep it simple' design framework for minimum metadata requirements.

Cases 4-6 are related to document-oriented searches, for which distribution format and online access are important, and a number of bibliographic properties (scale, publisher, series, series ID, media, file format) come into play.

3.4 Accessing resources

Explicit conventions for what URL's are included in metadata and how their behavior is specified are necessary to enable software to utilize them without operator intervention. Links in metadata to access resource should in general be complete URL's that can be invoked with a simple HTTP GET request, without having to add additional request parameters. Links should be accompanied by formal elements (with controlled vocabulary content) that provide machine-processable information to distinguish links that will return a document from links that invoke a service or access an online interactive application. The idea is that sufficient information should be provided that client software can parse the metadata record and provide useful functionality on the resource with minimal user interaction.

For many resources, different representations may be available. These might be different file formats for the same document for information resources. For non-information resources, a variety of representations that have different uses might be available. For example a physical sample may be represented by a text description of the sample, a GeoSciML xml description, visible light photograph, or images of the sample using other sensors. A geologic map may be available as a paper copy, a scanned image, a georeferenced scanned image, a vector data set in one of several formats (gml, shape file, file geodatabase, MIF, DWG), through a web map service, or through a web feature service. Metadata for a resource should be able to describe all of these different representations that the resource provider wishes to make available, in such a way that automated clients can seek representations useful to that client, or search clients can present users with links to access different formats or representations.

3.5 Citation and contact information

Citation information specifies the source of some content. Citations for the described resource specify the source for the resource intellectual content. The cited agent may have played various roles relative to the resource—author, compiler, editor, collector etc., and a controlled vocabulary is necessary to specify these. Citation for a metadata record specifies the agent responsible for producing the record, typically thought of as the metadata record creator. Metadata production involves elements of authoring, compiling, and editing. Minimally, citations must identify an individual person, an organization, or a role in an organization that is the agent filling a specified role relative to the cited resource. In most cases an organization will be specified, either as the employer or sponsor of a person, an institutional actor, or the host for some role (web master, metadata editor). In addition, information required to contact the cited actor is required to enable metadata users to contact a person with some knowledge of the cited resource. For long-lived metadata, contact for an agency role is most likely to persist. The minimum metadata contact information required is either an e-mail address or telephone number.

3.6 Fitness for purpose

The metadata should provide sufficient description of the resource for a user to determine if the resource is likely to meet their needs, and to determine what representation to access. At the simplest level, such information should be provided in the abstract in the metadata record. This puts the onus on metadata producers to document in the abstract information that will be useful for users to determine fitness. Such information includes why the resource was produced, what sort of observation procedures were used, assessment of data completeness, accuracy, and precision, and comparison with other known similar resources. The data quality section of ISO 19115 provides a data structure to formally describe this information, but the cost of using this is high (complex data entry), and there do not currently appear to be clients that utilize the information. The guiding principal should be that if users need to search on some particular quality criteria, specific guidance on how to encode that criteria in the metadata (which ISO 19139 elements, what controlled vocabulary to use if terminology is involved) is necessary. This is out of scope for a minimum metadata requirement.

3.7 Branding

In a distributed, federated catalog system with harvesting, metadata records are expected to propagate far beyond their original point of introduction into the system. If an organization producing metadata wishes to be recognized, and in order for users to be able to contact the metadata originator, contact information for the metadata originator must be considered part of the metadata record, and maintained in harvest processes. For presentation to users, it is desirable to provide a link to an icon that can be displayed with records to brand the origination of the metadata.

The same considerations hold for the resource itself.

3.8 Access constraints, legal limitations

Metadata records that are not for public consumption should never be exposed to a harvesting request. Implementation of security and access control must occur at a lower layer in the network stack than the catalog service is operating, such that authorization/ user authentication information is handled by the environment containing the catalog client and server. Metadata for commercially licensed resources may be publicly accessible, but should clearly indicate the licensing requirements and procedure.

3.9 Low cost of entry

Metadata producers should be able to reuse and build on existing structured metadata. Minimum requirements should be limited to information that is commonly available. Resource specific details should be provided in text elements in the metadata. Special information necessary to utilize web links (e.g. web service operation) in metadata should be provided by text in the metadata or through linked documents.

4 Content specification

Based on the above discussion, the following metadata content requirements are specified.

We are not proscribing any particular metadata format, but strongly recommend ISO19139 XML or FGDC CSDGM XML. Please confer with the AZGS developer team about metadata formatting to facilitate import of metadata into the NGDS catalog.

Explanation of fonts used: *Terms in italics are groupings of metadata properties*; **required**, **conditional**, and **optional** metadata content; (number of values that can be specified are in gray).

4.1 Minimum content

The follow list includes the minimum required content for basic resource description, discovery, and access. Several of the use case scenarios outlined above could not be supported with only this content.

- **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.

4.2 Recommended metadata content

This section extends the minimum content requirements with recommended content to produce useful metadata to describe resources, credit the originator of the resource, and inform users how to obtain or access a resource. The resource description should provide sufficient information to assist in discovery of the resource through an online search, and to allow users to evaluate the fitness of the resource for an intended purpose.

Explanation of fonts used: *Terms in italics are groupings of metadata properties*; **required**, **conditional**, and **optional** metadata content; (number of values that can be specified are in gray).

- *Resource*
  - **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): **North Bounding Latitude, South Bounding or Point Latitude, East Bounding Longitude, West Bounding or Point Longitude.** Values given in decimal degrees using the WGS 84 datum. Some resources may not be usefully described by an extent; if no extent is specified the default is Earth. This convention would have to be modified for systems describing extraterrestrial resources. For resources for which the geographic extent is not applicable, the recommendation is to include the keyword 'nonGeographic' in the subject (spatial) keywords element. If a particular encoding scheme requires a bounding box, a very small bounding rectangle will be created if only a point coordinate is given; the asserted point will be the southwest corner of the bounding box.
  - *Contact - Author or Intellectual Originator* (0 to 1 entry): The primary party responsible for creating the resource. **Organization Name**, **Person Name, Street Address**, **City**, **State**, **ZIP Code**, **Email**, **Phone**, **Fax, URL**. If contact information is provided, include at least the organization or author name.
  - **Bibliographic Citation** (0 to 1 entry): Full bibliographic citation if the resource has been published.
  - **Subject Keywords** (0 to many entries): Thematic, spatial and temporal free-form subject descriptors for the resource. A keyword may be assigned on metadata import if none are present. If possible, submit keywords in separate Thematic, Spatial, and Temporal keyword categories.
  - **Resource Language** (0 to 1 entry): Use three letter ISO 639-2 language code (defaults to "eng" for English).
  - **Resource ID** (0 to many entries): Resource identifier(s) following any public or institutional standard. Identified consists of an identifier string and if applicable a **Resource ID Protocol** identifier string that specifies the protocol for the resource ID standard. For example: undefined, ISBN-10, ISBN-13, ISSN, URN, URI, IRI, DOI, HTTP, SSN, etc.
    Examples: `doi:10.1000/182`; `isbn:0-671-62964-6`; `issn:1935-6862`; `azgs:OFR-10-02`
    Many protocols build the identifier for the protocol into the identifier string.
  - *Geographic Extent – Vertical* (0 to 1 entry*): **Datum Elevation**, **Upper Boundary, Lower Boundary, CRS**. The upper and lower boundaries of the extent are reported relative to the datum elevation. This will typically be the Earth surface at the location of the resource, or mean sea level (MSL). The coordinate reference system CRS value (normally a URI) specifies the datum for vertical measurement and the conventions used to define the coordinates of the extent boundaries relative to the datum (positive up, positive down, units of measurement). The upper boundary is farther from the center of the earth than the lower boundary.

*The same vertical extent may be reported using different coordinate reference systems in the same record to accommodate different conventions for searching. A borehole with collar elevation 4787 feet above sea level has core extracted at depths between 100 and 470 feet. The vertical extent of the core could be reported in either of the following ways: {0, CRS datum MSL positive up meters, 1420, 1308} or {1450.6, CRS datum Earth surface positive up meters, -30.3, -142.4}.

- *Temporal Extent* – Temporal range over which the resource was collected or is valid. If the resource pertains to specific named geologic time periods, those terms should be entered as keywords (preferable as part of Temporal Keywords). **Start Date** (0 to 1 entry), **End Date** (0 to 1 entry; required if start date exists),use ISO 8601 date and time format: YYYY-MM-DDTHH:MM:SS.
- **Quality Statement** (0 to 1 entry): Text specification of the quality of the resource.
- **Lineage Statement** (0 to 1 entry): Text description of the resource's provenance.

- *Access*
  - **Access Statement** (1 entry): Text instructions for how to access the resource.
  - *Distribution Contact* (1 entry): The party to contact about accessing the resource. **Organization Name**, **Person Name, Street Address**, **City**, **State**, **ZIP Code**, **Email**, **Phone**, **Fax, URL**. In general, a contact for distribution should be required for physical resources.
  - *Link to the resource* (0 to many entries): A URL pointing to a resource or resource webpage. **URL**, **Link Function**, **Representation Format**. URL is minimum content required if a link is included. A Link Function term from a controlled vocabulary to specify what a HTTP GET using the URL will invoke is recommended. The link might return an html page, electronic document in some other format, an end point for a service, an online application that requires user interaction, etc. Representation Format is a controlled vocabulary term specifying the format (MIME media types) of a file-based response if applicable.
  - **Constraints Statement** (0 to 1 entry): describe the resource's legal and usage constraints.
  - **Distribution Keywords** (0 to many entries): keywords describing the physical form of the resource (core, rock sample, digital file, book, journal article), formatting of resource content (file format, e.g. tiff, xls, MIME type), or physical distribution media (film, floppy disk, online service, hard copy). Table 6 in USGIN ISO metadata profile includes a vocabulary for distribution format for use with the ISO19115 distributionFormat name property. Use of these keywords allows users to search for particular kinds of artifacts.
- *Metadata*
  - **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* (1 entry): The party to contact with questions about the metadata itself. **Organization Name**, **Person Name, Street Address**, **City**, **State**, **ZIP Code**, **Email**, **Phone**, **Fax, URL**.
  - **Metadata Specification** (1 entry): Identifier string for the specification to which metadata record encoding is declared to conform. Should indicate the base standard and version, as well as any profile that applies to the content or encoding. Ideally the identifier could be dereferenced to obtain information about the applicable specification. Identifiers for metadata encoding specifications to be used in the USGIN and NGDS systems will have to be formally defined and registered for such identifiers to be broadly useful.
  - **Metadata UUID** (0 to 1 entry): A Universally Unique Identifier (UUID) will be assigned during the metadata import process if one is not provided. Unique identification of each metadata record is included to reduce duplication of entries across multiple metadata catalogs. The UUID format provides unique identification without centralized coordination.

4.2.1 Information that will be assumed unless specified otherwise

1. Character encoding of the metadata. Default is UTF-8.
2. Language of metadata (English)
3. Language of resource (English)

4.2.2 Resource specific requirements

1. Vertical extent is required for resources that pertain to a subsurface, ocean, or atmosphere location. If no vertical extent is specified, it is assumed to be the current Earth surface.
2. Published documents require a standard bibliographic citation (author, year, publisher, series, volume, page numbers, etc.) as specified by a publication style or guideline. Some example guidelines include USGS Suggestions to Authors and MLA Style Manual; the community will need to agree on conventions to use for citation syntax to improve interoperability. In general, for web-accessible digital resources that are the typical items of interest that will be cited, full text searches are anticipated to be the most common use case. Unless clear examples of use cases requiring more disaggregated representation of citations in the metadata (e.g. separate attributes for publisher, larger work title, larger work editor, volume, issue number, etc…) we will stick to simple text blob citations.
3. Spatial data specification require information on spatial resolution and terms to categorize spatial representation type: raster (spatial array), polygon, lines, and points
4. Web Services require:
    a. service type from controlled vocabulary. See Table 11 in USGIN ISO metadata profile for a starting-point interim vocabulary.
    b. URL for service-specific document that describes operation of service (e.g. OGC GetCapabilities, WSDL)
    c. Base URL for service requests
    d. Contact information for service provider

4.2.3 Optional but highly recommended

1. Citations for resource creator and metadata creator should include URL for icons to display to brand content in presentation to user.
2. Use scoped keywords from community thesauri to increase search efficiency. A gazetteer thesaurus like USGS place names is one obvious candidate. Details need to be determined.

4.3 Issues

How to deal with binding between related resources, like core association with logs from same borehole.

5 References

5.1 Cited literature

[Dublin Core] 2008-01-14 Dublin core Metadata Element Set, Version 1.1: Dublin Core Metadata Initiative, accessed at http://dublincore.org/documents/dces/.

Franklin, Michael, Halevy, Alon, and Maier, David, 2005, From databases to dataspaces: a new abstraction for information management: ACM SIGMOD Record, V. 34, No. 4, ISSN:0163-5808.

[ANZLIC, 2007] ANZLIC Metadata Profile Guidelines, Version 1.0: Turner, ACT, ANZLIC - the Spatial Information Council, ISBN: 978-0-646-46940-9, 372 p.

[INSPIRE ISO19115/119] Drafting Team Metadata and European Commision Joint Research Centre, 2009-02-18, INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119,v. 1.1: European Commission Joint Research Centre, MD_IR_and_ISO_20090218.

[USGIN 2010] 2010-0304, Use of ISO 19139 xml schema to describe geoscience information resources, v. 1.1.2, USGIN Standards and protocols drafting team, document gin2010-009, accessed at http://lab.usgin.org/sites/default/files/profile/file/u1/USGIN_ISO_Metadata_1.1.2.pdf.