USGIN DRAFT SPECIFICATIONS FOR GEOSCIENCE RESOURCE METADATA AND IDENTIFIERS

by Stephen Richard, Wolfgang Grunberg, and Ryan Clark, US Geoscience Information Network, Arizona Geological Survey, Tucson, Arizona
Geological Society of America, National Meeting, Nov. 1, 2010, Denver, Colorado

The catalog: a key system component in which data providers register and describe for discovery and access by users.

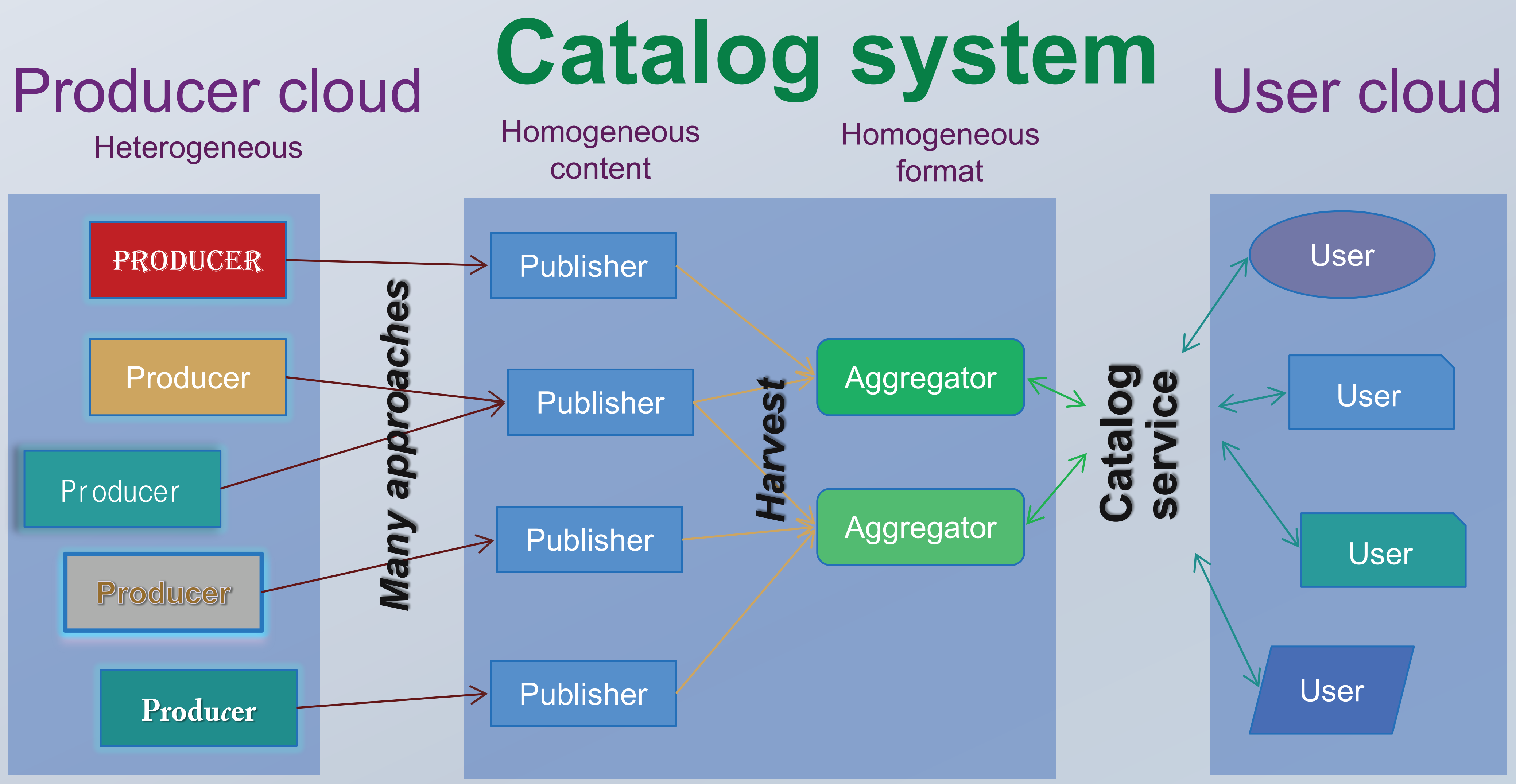
The metadata offered by the catalog must
-- guide search engines to locate requested resources
-- describe the resource content, provenance, and quality so users can determine if the resource will serve for intended usage
-- enable the user to obtain or access the resource
-- enable software to automatically access the resource

The USGIN project has been working on specifications to reduce the heterogeneity of 'standard' metadata and service implementations such that a single client can search against different catalogs, and successfully use the results. See **http://lab.usgin.org/profiles/usgin-iso-metadata-profile** for more information.

Discovery of resources across a wide spectrum of related geoscience disciplines requires a metadata content model with concepts that span those disciplines.

Specification-- instructions on how to do something
Standard-- a specification that alot of people use

Summary of USGIN metadata content recommendations.
See **http://lab.usgin.org/profiles/doc/metadata-content-recommendations** for complete discussion.



producers announce what they've got

Cross-domain generalized metadata and specialized discipline specific descriptions can use same pattern with different content models

Users search using many different applications

Resource

- o **Title** (1 entry): Succinct (preferably <250 characters) name of the resource.
- o **Description** (1 entry): Information about the resource's content as well as its context.
- o **Originators** (1 to many entries): Authors, editors, or corporate authors/curators of the resource.
- o **Publication Date** (1 entry): Publication, origination, or update date (not temporal extent) for the resource. Use ISO 8601 date and time format. 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.
- o **Geographic Extent - Horizontal** (1 entry, point or minimum bounding rectangle): North and South Bounding Latitude or Point Latitude, East and West Bounding Longitude 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. If a particular encoding scheme requires a bounding box, a minimum bounding rectangle will be created if only a point coordinates is given.
- o **Author or Intellectual Originator** (0 to 1 entry): The primary party responsible for creating the resource. Organization Name, Person Name, Position Name; contact point (Telephone, E-mail, postal).
- o **Bibliographic Citation** (0 to 1 entry): Full bibliographic citation if the resource has been published.
- o **Subject Keywords** (0 to many entries): Thematic, spatial and temporal free-form subject descriptors for the resource. If possible, submit keywords in separate Thematic, Spatial, and Temporal keyword categories.
- o **Resource Language** (0 to 1 entry): Language of resource content. Use three letter ISO 639-2 language code (defaults to "eng" for English).
- o **Resource ID** (0 to many entries): Resource identifier(s) following any public or institutional standard. Identifier 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.
- o **Geographic Extent – Vertical** (0 to 1 entry): Compound content includes Datum Elevation, Datum Type, Maximum Elevation, Minimum Elevation. Maximum and Minimum Elevations are relative to the reported datum elevation, which will typically be the Earth surface at the location of the resource or sea level. Datum Elevation must be reported relative to mean sea level (MSL) in meters using EPSG::5714 geodetic parameters (WGS 84). Datum type must be a controlled vocabulary (Earth surface, MSL, Kelly bushing, etc.). The maximum is always numerically greater than the minimum elevation. For boreholes with datum at the earth surface, depth below surface is reported as a negative number. Example: core from borehole at depths between 100 and 470 feet, borehole collar at 4787 feet above sea level. Vertical extent could be reported in either of the following ways: {0, "MSL", 1420, 1308} or {1450.6, "Earth surface", -30.3, -142.4}.
- o **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 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.
- o **Quality Statement** (0 to 1 entry): Text specification of the quality of the resource.
- o **Lineage Statement** (0 to 1 entry): Text description of the resource's provenance. Complex discipline-specific lineage (e.g. analytical procedures) are described in free text, and may be supplemented by discipline-specific content models.

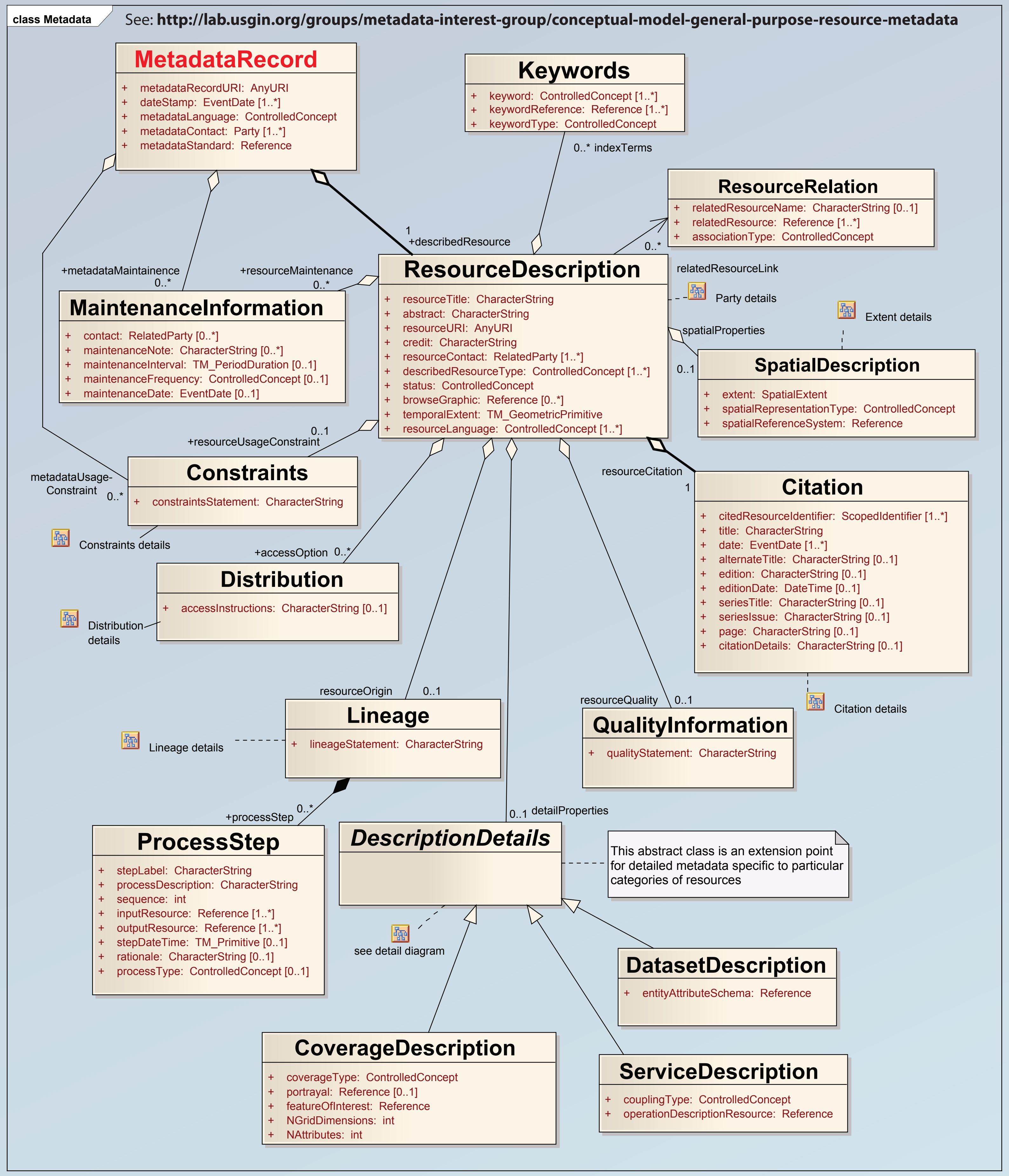
Access

- o **Access Statement** (1 entry): Text instructions for how to access the resource.
- o **Distribution Contact** (1 entry): Party to contact about accessing the resource. Organization Name, Person Name, Position Name; contact point either telephone, e-mail, or postal address. In general, a contact for distribution should be required for physical resources.
- o **Link to resource** (0 to many entries): A URL pointing to a resource or resource webpage. Element content includes URL, Link Function, and Representation Format. URL is minimum content required if a link is included. Link Function term from controlled vocabulary specifies what a HTTP GET using the URL will invoke. 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.
- o **Constraints Statement** (0 to 1 entry): describe the resource's legal and usage constraints in free text.
- o **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). Use of these keywords allows users to search for particular kinds of artifacts.

Metadata

- o **Metadata Date** (1 entry): Last metadata update/creation date-time stamp in ISO 8601 date and time format.
- o **Metadata Contact** (1 entry): The party to contact with questions about the metadata itself. Organization Name, Person Name, Position Name; contact point either telephone, e-mail, or postal address.
- o **Metadata Specification** (1 entry): Identifier string for the metadata specification used to create the metadata record encoding this content. 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 will have to be formally defined and registered for such identifiers to be broadly useful.
- o **Metadata UUID** (0 to 1 entry): A Universally Unique Identifier (UUID) for the metadata record; required to avoid duplicate entries across multiple metadata catalogs. The UUID format provides unique identification without centralized coordination.

Metadata model


class Metadata
See: http://lab.usgin.org/groups/metadata-interest-group/conceptual-model-general-purpose-resource-metadata
MetadataRecord
+ metadataRecordURI: AnyURI
+ dateStamp: EventDate [1..*]
+ metadataLanguage: ControlledConcept
+ metadataContact: Party [1..*]
+ metadataStandard: Reference
Keywords
+ keyword: ControlledConcept [1..*]
+ keywordReference: Reference [1..*]
+ keywordType: ControlledConcept
0..* indexTerms
ResourceRelation
+ relatedResourceName: CharacterString [0..1]
+ relatedResource: Reference [1..*]
+ associationType: ControlledConcept
0..*
relatedResourceLink
1
+describedResource
+metadataMaintainence
0..*
+resourceMaintenance
0..*
ResourceDescription
+ resourceTitle: CharacterString
+ abstract: CharacterString
+ resourceURI: AnyURI
+ credit: CharacterString
+ resourceContact: RelatedParty [1..*]
+ describedResourceType: ControlledConcept [1..*]
+ status: ControlledConcept
+ browseGraphic: Reference [0..*]
+ temporalExtent: TM_GeometricPrimitive
+ resourceLanguage: ControlledConcept [1..*]
Party details
Extent details
spatialProperties
0..1
SpatialDescription
+ extent: SpatialExtent
+ spatialRepresentationType: ControlledConcept
+ spatialReferenceSystem: Reference
MaintenanceInformation
+ contact: RelatedParty [0..*]
+ maintenanceNote: CharacterString [0..*]
+ maintenanceInterval: TM_PeriodDuration [0..1]
+ maintenanceFrequency: ControlledConcept [0..1]
+ maintenanceDate: EventDate [0..1]
0..1
+resourceUsageConstraint
metadataUsage-Constraint
0..*
Constraints
+ constraintsStatement: CharacterString
Constraints details
+accessOption 0..*
Distribution
+ accessInstructions: CharacterString [0..1]
Distribution details
resourceCitation
1
Citation
+ citedResourceIdentifier: ScopedIdentifier [1..*]
+ title: CharacterString
+ date: EventDate [1..*]
+ alternateTitle: CharacterString [0..1]
+ edition: CharacterString [0..1]
+ editionDate: DateTime [0..1]
+ seriesTitle: CharacterString [0..1]
+ seriesIssue: CharacterString [0..1]
+ page: CharacterString [0..1]
+ citationDetails: CharacterString [0..1]
Citation details
resourceOrigin 0..1
Lineage
+ lineageStatement: CharacterString
Lineage details
resourceQuality 0..1
QualityInformation
+ qualityStatement: CharacterString
+processStep 0..*
ProcessStep
+ stepLabel: CharacterString
+ processDescription: CharacterString
+ sequence: int
+ inputResource: Reference [1..*]
+ outputResource: Reference [1..*]
+ stepDateTime: TM_Primitive [0..1]
+ rationale: CharacterString [0..1]
+ processType: ControlledConcept [0..1]
0..1 detailProperties
DescriptionDetails
This abstract class is an extension point for detailed metadata specific to particular categories of resources
see detail diagram
DatasetDescription
+ entityAttributeSchema: Reference
CoverageDescription
+ coverageType: ControlledConcept
+ portrayal: Reference [0..1]
+ featureOfInterest: Reference
+ NGridDimensions: int
+ NAttributes: int
ServiceDescription
+ couplingType: ControlledConcept
+ operationDescriptionResource: Reference

Based on ISO 19115 and FGDC content.

Modified to be explicit while avoiding implementation-specific artifacts where possible

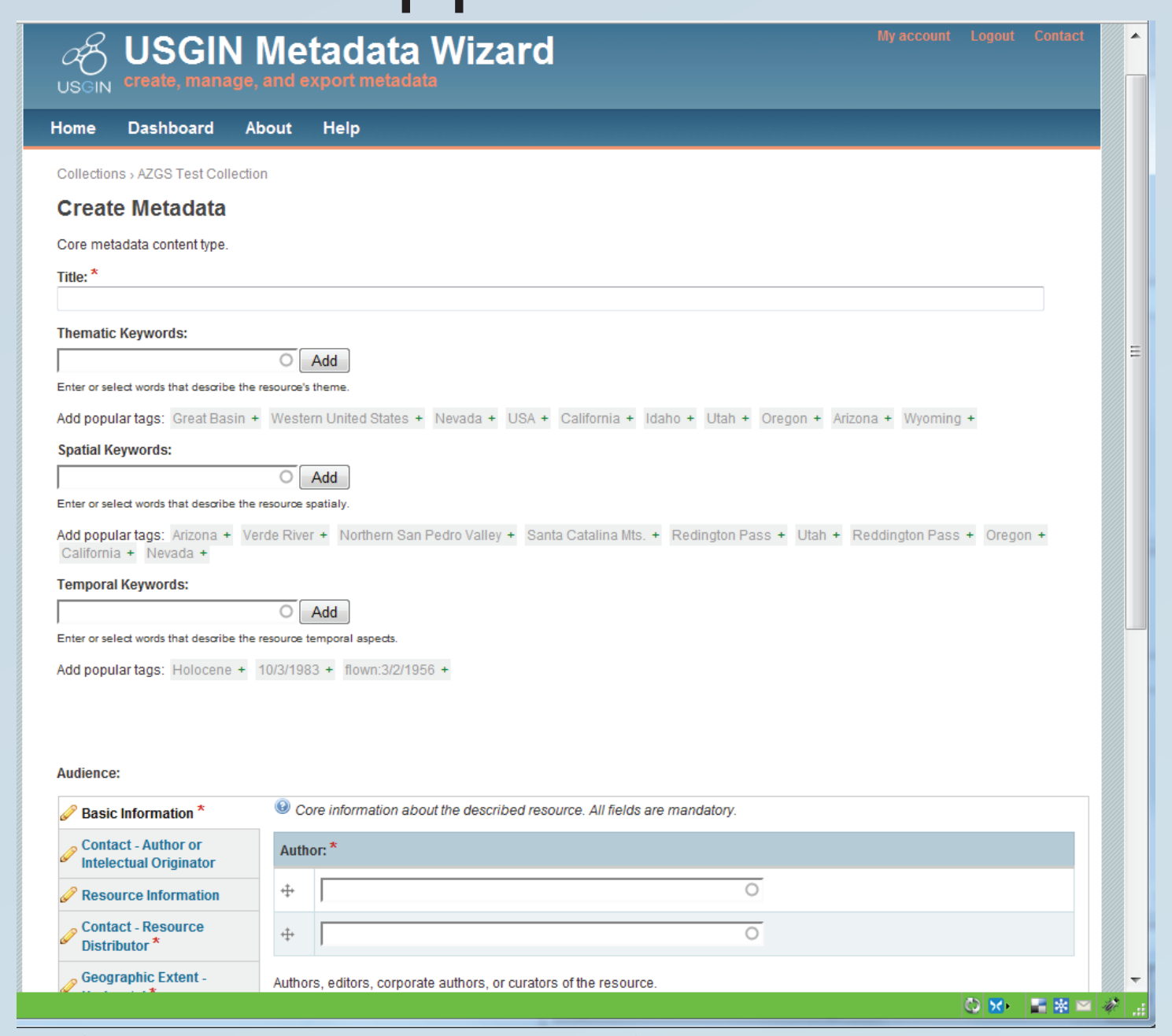
Focus is on attributes that will apply to broad spectrum of resources

Discipline-specific details are implemented by extensions from DescriptionDetails abstract class

Various details of associated with the various base attribute classes are not shown in this diagram. See http://lab.usgin.org/groups/ metadata-interest-group for the full model UML.

Implementation of this model as an ontology and mapping of various schemes in use to a single model would enable better metadata interoperability.

Applications



Collections › AZGS Test Collection

Create Metadata

Audience:

Basic Information *
Core information about the described resource. All fields are mandatory.

Contact - Author or Intelectual Originator
Author: *

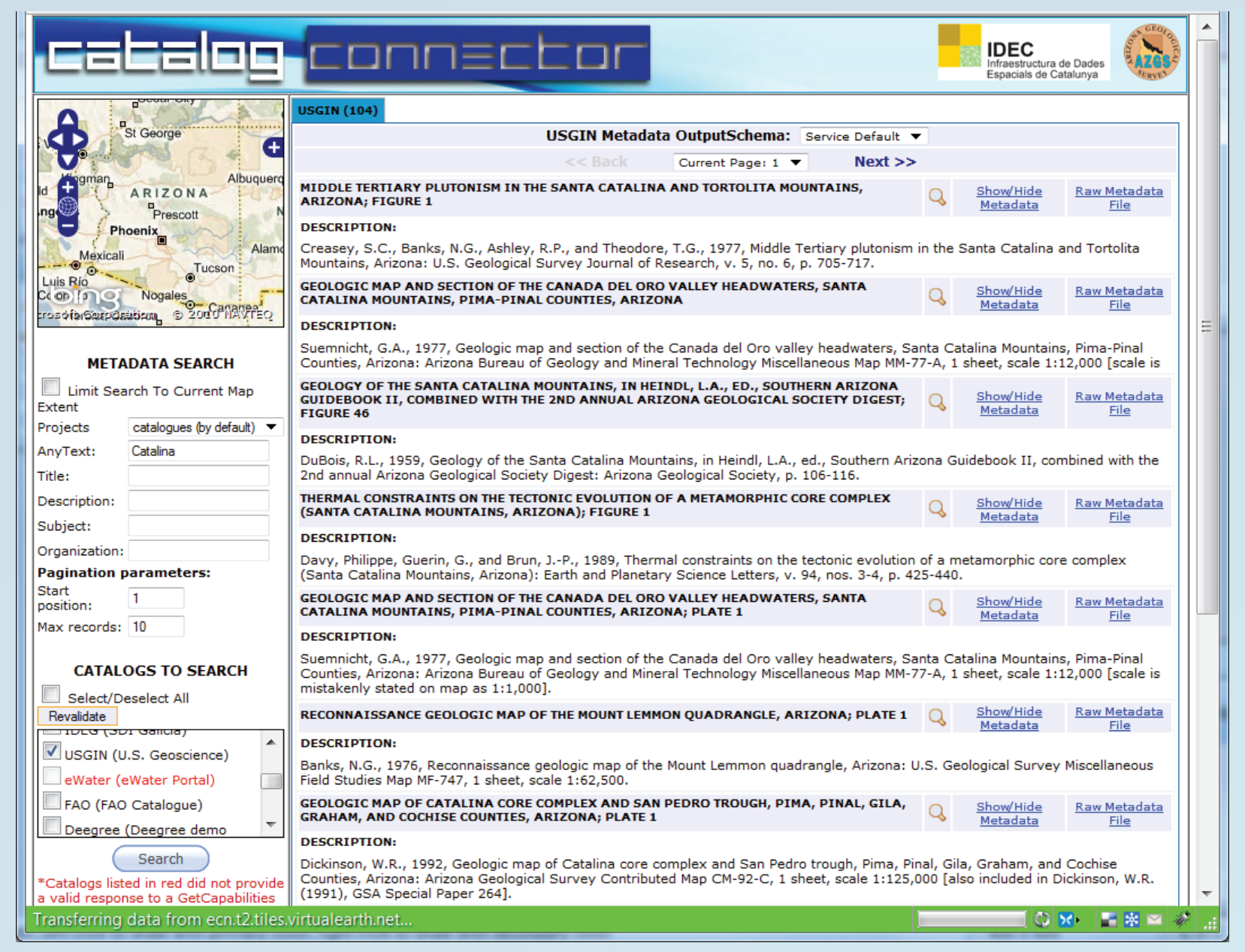
Resource Information

Contact - Resource Distributor *

Geographic Extent -

Authors, editors, corporate authors, or curators of the resource.

Online tool for generating metadata in USGIN profile; will export FGDC xml or ISO19139 xml. Access at http://mw.usgin.org



USGIN Metadata OutputSchema: Service Default
<< Back Current Page: 1 Next >>

METADATA SEARCH
□ Limit Search To Current Map Extent
Projects: catalogues (by default)
AnyText: Catalina
Title:
Description:
Subject:
Organization:
Pagination parameters:
Start position: 1
Max records: 10

CATALOGS TO SEARCH
□ Select/Deselect All
☑ USGIN (U.S. Geoscience)
□ eWater (eWater Portal)
□ FAO (FAO Catalogue)
□ Deegree (Deegree demo
*Catalogs listed in red did not provide a valid response to a GetCapabilities

MIDDLE TERTIARY PLUTONISM IN THE SANTA CATALINA AND TORTOLITA MOUNTAINS, ARIZONA; FIGURE 1 — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
Creasey, S.C., Banks, N.G., Ashley, R.P., and Theodore, T.G., 1977, Middle Tertiary plutonism in the Santa Catalina and Tortolita Mountains, Arizona: U.S. Geological Survey Journal of Research, v. 5, no. 6, p. 705-717.

GEOLOGIC MAP AND SECTION OF THE CANADA DEL ORO VALLEY HEADWATERS, SANTA CATALINA MOUNTAINS, PIMA-PINAL COUNTIES, ARIZONA — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
Suemnicht, G.A., 1977, Geologic map and section of the Canada del Oro valley headwaters, Santa Catalina Mountains, Pima-Pinal Counties, Arizona: Arizona Bureau of Geology and Mineral Technology Miscellaneous Map MM-77-A, 1 sheet, scale 1:12,000 [scale is

GEOLOGY OF THE SANTA CATALINA MOUNTAINS, IN HEINDL, L.A., ED., SOUTHERN ARIZONA GUIDEBOOK II, COMBINED WITH THE 2ND ANNUAL ARIZONA GEOLOGICAL SOCIETY DIGEST; FIGURE 46 — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
DuBois, R.L., 1959, Geology of the Santa Catalina Mountains, in Heindl, L.A., ed., Southern Arizona Guidebook II, combined with the 2nd annual Arizona Geological Society Digest: Arizona Geological Society, p. 106-116.

THERMAL CONSTRAINTS ON THE TECTONIC EVOLUTION OF A METAMORPHIC CORE COMPLEX (SANTA CATALINA MOUNTAINS, ARIZONA); FIGURE 1 — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
Davy, Philippe, Guerin, G., and Brun, J.-P., 1989, Thermal constraints on the tectonic evolution of a metamorphic core complex (Santa Catalina Mountains, Arizona): Earth and Planetary Science Letters, v. 94, nos. 3-4, p. 425-440.

GEOLOGIC MAP AND SECTION OF THE CANADA DEL ORO VALLEY HEADWATERS, SANTA CATALINA MOUNTAINS, PIMA-PINAL COUNTIES, ARIZONA; PLATE 1 — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
Suemnicht, G.A., 1977, Geologic map and section of the Canada del Oro valley headwaters, Santa Catalina Mountains, Pima-Pinal Counties, Arizona: Arizona Bureau of Geology and Mineral Technology Miscellaneous Map MM-77-A, 1 sheet, scale 1:12,000 [scale is mistakenly stated on map as 1:1,000].

RECONNAISSANCE GEOLOGIC MAP OF THE MOUNT LEMMON QUADRANGLE, ARIZONA; PLATE 1 — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
Banks, N.G., 1976, Reconnaissance geologic map of the Mount Lemmon quadrangle, Arizona: U.S. Geological Survey Miscellaneous Field Studies Map MF-747, 1 sheet, scale 1:62,500.

GEOLOGIC MAP OF CATALINA CORE COMPLEX AND SAN PEDRO TROUGH, PIMA, PINAL, GILA, GRAHAM, AND COCHISE COUNTIES, ARIZONA; PLATE 1 — Show/Hide Metadata — Raw Metadata File
DESCRIPTION:
Dickinson, W.R., 1992, Geologic map of Catalina core complex and San Pedro trough, Pima, Pinal, Gila, Graham, and Cochise Counties, Arizona: Arizona Geological Survey Contributed Map CM-92-C, 1 sheet, scale 1:125,000 [also included in Dickinson, W.R. (1991), GSA Special Paper 264].

Transferring data from ecn.t2.tiles.virtualearth.net...

Light weight catalog client will search a variety of metadata catalogs. Code project started by IDEC (Zaragoza, Spain) with contributions from USGIN. http://catalog.usgin.org/catalogConnector

Resource Identifiers

Stable, dereferenceable identifiers are an essential component of a linked data system

Stable: mapping between identifier string and identified thing is immutable.

Dereferenceable: a well know system exists to present user with an explanation or representation of what is identified.

Resource type	notes	examples
concrete information resource	a specific representation of a resource. Equivalence is determined by bitwise matching of the representation bitstream	monaLisa.tif, USGIN_URI_scheme_1.0.1.doc
abstract information resource	an information resource defined by a particular content scope, includes most kinds of documents. Equivalence is determined by matching of content meaning or intention (depending on how the resource is defined)	Image of Mona Lisa, current version of USGIN URI scheme, the Declaration of Independence
concrete non-information resoruce	An identifiable physical object or event.	The original Mona Lisa painting, The original Declaration of Independence parchment, a particular rock sample, a particular car, a particular meeting, a particular project.
abstract non-information resource	A concept defined by some human intention, may categorize other resources, or represent some abstract idea.	painting, love, the US Congress

Kinds of resources that may be identifier by a URI

Dereferencing URI for information resource:

--- Present the normative (canonical form) actual resource.

Dereferencing URI for a non-information resource:

--- Present a representation of the resource. There may be many valid representations of non-information resources, useful in different contexts. Each of these is an identifiable information resource and the canonical or default representation may be context dependent.

URI Scheme (see **http://lab.usgin.org/profiles/doc/usgin-uri-recommendations-v10**)

"http:" "//" `uriHost` "/uri_gin/" `nameAuthority` "/" `resourcePath`
["/" `resourceSpecificString`] [("/" / "/" `representationPart`)]

`uriHost` -- host system on the internet that will de-reference the following identifier. Not considered part of the USGIN resource identifier for string comparisons of identifier equivalence, but usage of a known host conveys trust. Syntax follows standard web conventions, e.g. "azgs.az.gov" .

The sequence of characters between "uri_gin/" and the last "/" in the URI constitute the identifier for a resource.

`nameAuthority` Name of the naming authority for the identified resource, taken from the register with USGIN URI. A name authority defines the collection of resource categories for which it maintains identifiers.

`resourcePart` -- A path that identifies a resource category and specific instance of a resource of that type. Includes:

`resourcePath` Specifies the resource category, taken from the register at http://resources.usgin.org/uri_gin/usgin/register/resourceType/. Each resource type definition should specify what the default and canonical representations are (they may be the same or different), the known valid representations, and the dereferencing behavior for the resource type. The resourcePath is hierarchical to allow representation of resource hierarchy when that is useful.

`resourceSpecificString` A string that identifies a particular instance of the resource type in the resource path, unique in the scope of the name authority. This string may have syntax specially scoped for a particular resource type. Definition of resourceSpecificString syntax should be in the resource type definition.

`representationPart` is intended to identify a representation of the preceding resource part. Dereferencing behavior for such an identifier is defined in the associated resource type definition. Note that an abstract document is an information resource, so in this case the representationPart is semantically equivalent to a resourceSpecificString.

GIN is supported in part by the National Science Foundation under EAR - 0753154 to the Arizona Geological Survey acting on behalf of the Association of American State Geologists, and by the USGS National Geologic Map Database.

U.S. DEPARTMENT OF ENERGY

USGIN is supported by: The U.S. Department of Energy under award DE-EE0001120 to Boise State University and under award DE-EE1002850 to the Arizona Geological Survey acting on behalf of the Association of American State Geologists.

Arizona Geological Survey
416 W. Congress St. Suite 100
Tucson, AZ 85701