

Creating and Publishing Metadata in Support of Geospatial One-Stop and the NSDI

Background.....	1
Metadata Content.....	1
Going Beyond the Minimum	1
Multiple Online Linkages.....	2
Theme Keywords Using ISO Topic Categories.....	2
Metadata Creation Strategies.....	3
Create and Use Templates.....	3
Document Existing and Planned Data.....	3
Publishing Metadata via geodata.gov.....	3
Metadata Harvesting.....	3
What is Metadata Harvesting?	3
Requirements.....	4
How Harvesting Works	4
Types of Harvesting Protocols	5
Metadata Upload.....	6
Metadata Direct Entry.....	6
Appendices	
A. Theme Keywords and the ISO Topic Categories.....	7
B. Required FGDC Tags and Validation Rules.....	10
C. Sample XML Metadata Record with GOS Essential Elements	12

Background

Metadata has long been promoted by the Federal Geographic Data Committee (FGDC), state GIS coordinators, and project managers as a means to:

- preserve organizational data investments
- instill data accountability and liability, and
- facilitate data sharing.

With the implementation of the Geospatial One-Stop (GOS) federal E-gov initiative, metadata is established as the official language for national data development and exchange. To maximize the value of your metadata investment and become an active participant in the GOS Geodata.gov portal, consider the following:

Metadata Content

Going Beyond the Minimum

Minimal metadata is minimally useful. If you limit your metadata to the mandatory elements of the Content Standard for Digital Geospatial Metadata (CSDGM), then you have limited your metadata to those elements common to all data types and have not realized the value of metadata to capture that which is unique to the data set. Metadata producers must determine all metadata elements necessary to adequately characterize the data set and provide complete, current information for each element. When you create metadata that goes beyond the minimum, you create a data management resource that serves both your community and your own data management efforts.

Multiple Online_Linkage Values

As a 'repeatable' element, **Online_Linkage** (Citation Information) is used to provide access to a variety of data download, data clearinghouse, and web-mapping services. Use this field to fully represent your geospatial data access and distribution capabilities by providing complete URLs and necessary information to indicate the nature of the weblink using the following style guidance:

- OGC Web Map Service (WMS) links include a 'getmap' request with a layer name, version, preferred image format, and preferred SRS, at a minimum: <http://server/service?REQUEST=getmap&VERSION=1.1.0&LAYERS=roads&FORMAT=image/gif&SRS=EPSG:4326>
- ArcIMS "Image" services using a URL-like request. If you pasted this request in a browser you will not connect to an ArcIMS server since it does not permit this style of request, however it contains enough information to allow geodata.gov to connect to an ArcIMS service: http://<server>/image/<service_name> will be assigned as Live Map, ArcXML Image service, where server URL is <server> and service name is ArcIMS <service_name>. The sub-path "/image/" must be present in the URL
- direct download sites include URLs, that start with either ftp:// or http:// and point to filenames with .zip, .tar, .tgz, .gz, .dxf, or .e00 extensions.

Theme_Keywords Using ISO Topic Categories

The more robust your theme keyword list, the more likely it can be located by others (and yourself). Data discovery is further enabled by the use of standardized keyword lists and vocabularies (Theme_Keyword_Thesaurus). The FGDC and GOS utilize the International Organization for Standards Metadata Standard (ISO 19115) Topic Categories to organize data and services into *data categories*. A list of Topic Categories with definitions and examples is provided on page 7 of this document. When creating new metadata records, include one or more Topic Categories as Theme_Keywords and cite the Theme_Keyword_Thesaurus as 'ISO 19115 Topic Categories'. Use the exact format of the Topic Category values (e.g. "biota", see Appendix A for more detail) unless your metadata creation tool provides a pick list of ISO-based themes.

For existing metadata records, develop a strategy for incorporating one or more Topic Categories into each metadata record. For homogenous metadata collections (those that contain metadata records related to a single Topic Category term) a simple script can be written to insert the Topic Category term into the Theme_Keyword element of each record. For heterogenous metadata collections (those that contain metadata records relating to a variety of subjects) the task will be more challenging. For collections with few metadata records, the Topic Category terms can be inserted manually into the Theme_Keyword element. For more extensive collections, all existing Theme_Keywords can be output to a listing and, as suggested above, a lookup table developed that relates a set of Theme_Keywords to each applicable Topic Category. Once the lookup table is developed, a script can be written to insert all applicable Topic Categories into the Theme_Keyword element of each record.

If you are registering your metadata collection with geodata.gov to make your metadata available for harvesting into the geodata.gov portal, the metadata publisher registration process allows you to edit or associate a lookup table that will automatically assign an ISO Topic Category value when the validator encounters an equivalent term you have used in your Theme_Keyword field. **However**, if you use the geodata.gov publisher

registration option, the Topic Category term will not be added to your metadata record but simply used to direct your information to the geodata.gov channel dedicated to the associate Topic Category. While this enables your metadata to be better utilized within geodata.gov, you have lost the additional benefit of 'prepping' your records for easy translation to the eventual adoption of the international (ISO) metadata standard.

Metadata Creation Strategies

Create and Use Templates

Organizations are encouraged to create metadata templates that establish core content for all organizationally produced metadata. Templates should:

- outline all metadata elements deemed mandatory by the organization
- provide standardized language for access, use, and liability statements
- provide definitions and domains for standardized data layers
- establish standards and guidelines for metadata production and publishing.

Document Existing and *Planned* Data

In addition to documenting existing data resources NSDI participants are encouraged to use the CSDGM to document **planned data acquisitions**.

When creating metadata for planned data:

1. indicate the **Status** element as '*planned*'
2. provide a robust **Abstract**, including data type (vector, image, raster...), geographic location, and specifications (scale, film type, bands...)
3. include a rich set of **Theme / Place_Keywords**, including the ISO Topic Categories as Theme Keyword values, as described above

The documentation of planned data acquisitions enables developers to leverage data development investments via partnerships. In FY05, all federal agencies are required to create metadata for data acquisition plans estimated at \$500K or greater.

Publish Metadata via geodata.gov

To make your metadata records available via Geodata.gov you must either:

- publish your metadata collection to a metadata distribution server from which Geodata.gov can harvest,
- directly upload your XML formatted metadata to geodata.gov, or
- create your metadata online using the geodata.gov metadata publication tool.

Metadata Harvesting

What is metadata harvesting?

Metadata harvesting is an automated scheduled process for collecting new and updated metadata from a wide variety of GIS metadata sources. The process of harvesting allows geodata.gov to synchronize its metadata with publishers metadata. If you participate in metadata harvesting, any update to your metadata should be made on your metadata repository. geodata.gov will obtain the update through harvesting.

If you have registered on geodata.gov as a publisher and would like to participate in harvesting, you need to update your publisher information.

Currently geodata.gov can harvest FGDC-compliant metadata from four different type of harvesting protocols: (1) Z39.50 metadata clearinghouse node, (2) ArcIMS

metadata service, (3) Web Accessible Folder (WAF), and (4) Open Archive Initiative (OAI) metadata service.

Requirements

For a metadata record to be successfully harvested by geodata.gov, the following must be present in some form:

- **Document unique ID** - Document unique ID in each metadata is required to determine if a document is new to geodata.gov. If your metadata clearinghouse is a Z39.50 type, you need to verify if document unique ID has been implemented in each metadata document. For Isite, please check the Isite distribution to obtain the new release of Isite (Isite Vers. 2.10) which implements document unique ID and update date.
Document unique ID is not an issue if your repository is a WAF, OAI metadata service, or an ArcIMS metadata service because these services already handle and expose a unique identifier for each document in the collection.
- **Update date** - Once your repository has been harvested, the next harvesting will only look for metadata documents that are updated since the last harvesting date. In all cases the update date should be reflected in the "Metadata Date" field of CSDGM metadata.
- **Keywords** - Keywords are used to correctly categorize your metadata. geodata.gov uses standard theme keywords as specified in the ISO 19115. Without standard keywords, your metadata will still be published and searchable in the geodata.gov repository but it will not be categorized in one of the data categories.

You can submit standard keywords using one of the following methods:

- Insert the theme keyword with a standard keyword in the metadata
 - Provide a lookup table in the harvesting registration process that translates your localized keywords into standard keywords.
- **Register on geodata.gov** - Before your metadata repository can be harvested, you need to register as a publisher, read and accept the publisher disclaimer, specify the type of harvesting in the publishing registration form, and provide keywords.

How Harvesting Works

Metadata harvesting in geodata.gov is performed in three steps:

1. Harvesting – based on information provided during the registration, geodata.gov will connect to your metadata repository, retrieve all metadata records if it is the first time harvesting, or only the updated records since the last harvesting date. You need to verify that the date of creation or last update is stored in your metadata (Metadata Date).
2. Validation – during validation, each metadata record will be examined to meet the minimum requirements (see the list at the end of this description). The validation function will recognize only the FGDC tags. You can access your validation report via the harvesting history function.
3. Publishing – during publishing, all successfully validated metadata will be published in the geodata.gov. If the same document (as indicated by document unique ID) already exists, then the existing document will be

updated. Otherwise the document will be inserted in geodata.gov as a new document. Once the metadata is published, it will be searchable from geodata.gov.

4. Data Type Assignment – during validation, metadata records that include Online_Linkage values will be automatically assigned to a specific 'Data Type' based upon the URL provided.

Types of harvesting protocols

geodata.gov supports four types of harvesting protocol:

1. [Z39.50 Metadata Clearinghouse](http://www.fgdc.gov/clearinghouse/clearinghouse.html)
<http://www.fgdc.gov/clearinghouse/clearinghouse.html>
If your repository is a Z39.50 metadata clearinghouse, you need to verify whether or not a document unique ID and update date are implemented in each metadata document. For Isite, please check the Isite distribution location (<http://clearinghouse4.fgdc.gov/ftp>) to get the Version 2.10 release of Isite that implements document unique ID and update date. For SMMS GeoConnect, Blue Angel, Compusult MetaManager and other Z39.50 software providers, contact your distributor for information regarding unique ID and update date capabilities. To register your Z39.50 node to be harvested, you need to provide the URL, port number and database name.
Note: If your collection is less than 200 records, consider establishing a Web Accessible Folder (as described in Item 3. below) as an alternative to implementing the unique ID and update date features.
2. [ArcIMS Metadata Service](http://www.esri.com/software/arcims/overview.html)
<http://www.esri.com/software/arcims/overview.html>
If you currently maintain and serve your metadata using an ArcIMS metadata service, you will need to specify the URL, service name, and if applicable, the username and password to browse metadata.
3. Web Accessible Folder (WAF)
You can participate in geodata.gov harvesting by simply locating your metadata in XML files on a WAF. A WAF is a directory on the WWW where a Web browser can browse the content of the directory. It may not contain a default.html or index.htm file. To register your WAF to be harvested, you only need register the URL. It is recommended, but not required, that you also include html versions of the metadata records within the WAF to support discovery by search engines such as Google.
4. [Open Archive Initiative \(OAI\) Metadata Service](http://www.openarchives.org/)
<http://www.openarchives.org/>
If you maintain and serve your metadata using OAI Protocol for Metadata Harvesting, you need to provide the URL, set name and metadata prefix.

Metadata Upload

If you do not have access to any form of metadata distribution server as described above, you can upload your XML formatted metadata records directly to geodata.gov. You must first register at geodata.gov then select the 'Upload Metadata' option. Uploads can be done individually by record or as a group in batch mode. Please note that if you utilize the upload option, your published metadata is not linked to your resident metadata in any manner and updates to the metadata record must be uploaded manually as they cannot be automatically updated by geodata.gov.

Metadata Direct Entry

If you do not have access to metadata creation software/editor, or you have very few records to contribute to geodata.gov, you can utilize the online metadata creation tool provided at geodata.gov. As with the other metadata publishing options, you must first register as a metadata publisher at the geodata.gov site then select the 'Create Metadata' option. Note that the metadata will then be stored at geodata.gov and all updates must be made via geodata.gov. Also, the geodata.gov online metadata creation tool is intended as an easy to use interface for the collection of those metadata elements necessary for data discovery. As such metadata records created using the online tool will be limited in their use as data archive and data management resources.

Appendix A

Preparing for the international metadata standard: Theme Keywords and the ISO Topic Categories

The International Organization for Standards (ISO) metadata standard (ISO 19115) provides a set of Core metadata elements that must occur in every national profile/implementation. Most of these elements either map to existing CSDGM metadata elements or represent properties of the data that can be determined and populated using a data integrated metadata tool. *Topic Category* is the only mandatory element of the ISO core metadata set that requires new information that cannot be directly captured from the data. The following 19 subject headings represent the domain for the *Topic Category* element.

If your metadata creation software provides a pick list of Topic Category related terms simply select the pick list terms that apply and the software will insert the related Topic Category Name and/or Code. If creating data using the **Geodata.gov** metadata publisher, you will be asked to select a *Primary Theme*. The *Primary Theme* options are based upon the *ISO Topic Categories* below but the names have been altered to provide greater context, e.g., Geodata.gov Primary Theme 'Cultural, Society, and Demographic' will be captured in the Theme_Keyword metadata element as ISO Topic Category Name 'Society'.

If your metadata creation software does not provide a list of subject headings based upon the ISO 19115 Topic Category, include the *Topic Category Names* (as presented below) as *Theme_Keywords* and cite your related *Theme_Keyword_Thesaurus* as: 'ISO 19115 Topic Category'. The FGDC intends to develop CSDGM to ISO translation software that will insert the *Topic Category Code* when the *Topic Category Name* is found, however, those wishing to include the *Topic Category Code* as a *Theme_Keyword* can do so using the same *Theme_Keyword_Thesaurus*: 'ISO 19115 Topic Category'.

Include **all** pertinent Topic Category Names, e.g.,:
 business districts = *boundaries* and *economy*
 toxic release inventory = *environment* and *health*
 soil fertility = *geophysical* and *farming*

ISO Topic Category Name Code

farming	001	rearing of animals and/or cultivation of plants Examples: agriculture, irrigation, aquaculture, plantations, herding, pests and diseases affecting crops and livestock
biota	002	flora and/or fauna in natural environment Examples: wildlife, vegetation, biological sciences, ecology, wilderness, sea life, wetlands, habitat, biological resources
boundaries	003	legal land descriptions Examples: political and administrative boundaries, governmental units, marine boundaries, voting districts, school districts, international boundaries
climatology Meteorology Atmosphere	004	processes and phenomena of the atmosphere Examples: cloud cover, weather, climate, atmospheric conditions, climate change, precipitation

economy	005	economic activities, conditions, and employment Examples: production, labor, revenue, business, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, exploration and exploitation of resources such as minerals, oil and gas
elevation	006	height above or below seal level Examples: altitude, bathymetry, digital elevation models, slope, derived products, DEMs, TINs
environment	007	environmental resources, protection and conservation Examples: environmental pollution, waste storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape, water quality, air quality, environmental modeling
geoscientificInformation	008	information pertaining to earth sciences Examples: geophysical features and processes, geology, minerals, sciences dealing with the composition, structure and origin of the earth's rocks, risks of earthquakes, volcanic activity, landslides, gravity information, soils, permafrost, hydrogeology, groundwater, erosion
health	009	health, health services, human ecology, and safety Examples: disease and illness, factors affecting health, hygiene, substance abuse, mental and physical health, health services, health care providers, public health
imageryBaseMapsEarthCover	010	base maps Examples: land/earth cover, topographic maps, imagery, unclassified images, annotations, digital ortho imagery
intelligenceMilitary	011	military bases, structures, activities Examples: barracks, training grounds, military transportation, information collection
inlandWaters	012	inland water features, drainage systems and characteristics Examples: rivers and glaciers, salt lakes, water utilization plans, dams, currents, floods and flood hazards, water quality, hydrographic charts, watersheds, wetlands, hydrography
location	013	positional information and services Examples: addresses, geodetic networks, geodetic control points, postal zones and services, place names, geographic names
oceans	014	features and characteristics of salt water bodies (excluding inland waters) Examples: tides, tidal waves, coastal information, reefs, maritime, outer continental shelf submerged lands, shoreline
planningCadastre	015	information used for appropriate actions for future use of the land Examples: land use maps, zoning maps, cadastral surveys, land ownership, parcels, easements, tax maps, federal land ownership status, public land conveyance records

society	016	characteristics of society and culture Examples: settlements, housing, anthropology, archaeology, education, traditional beliefs, manners and customs, demographic data, tourism, recreational areas and activities, parks, recreational trails, historical sites, cultural resources, social impact assessments, crime and justice, law enforcement, census information, immigration, ethnicity
structure	017	man-made construction Examples: buildings, museums, churches, factories, housing, monuments, shops, towers, building footprints, architectural and structural plans
transportation	018	means and aids for conveying persons and/or goods Examples: roads, airports/airstrips, shipping routes, tunnels nautical charts, vehicle or vessel location, aeronautical charts, railways
utilitiesCommunication	019	energy, water and waste systems and communications infrastructure and services Examples: hydroelectricity, geothermal, solar and nuclear sources of energy, water purification and distribution, sewage collection and disposal, electricity and gas distribution, data communication, telecommunication, radio, communication networks

Appendix B

Required FGDC XML Tags and Validation Rules

- Data Originator
Tag: /metadata/idinfo/citation/citeinfo/origin/
Rule: not null
Domain: "unknown" or free text
- Data Title
Tag: /metadata/idinfo/citation/citeinfo/title/
Rule: not null
Domain: free text
- Abstract
Tag: /metadata/idinfo/descript/abstract/
Rule: not null
Domain: free text
- Progress
Tag: /metadata/idinfo/status/progress
Rule: not null
Domain: "complete", "in work", "planned"
- West Bounding Coordinate
Tag: /metadata/idinfo/spdom/bounding/westbc/
Rule: not null
Domain: number between (-180.00) and (180.00)
- East Bounding Coordinate
Tag: /metadata/idinfo/spdom/bounding/eastbc/
Rule: not null
Domain: number between (-180.00) and (180.00)
- North Bounding Coordinate
Tag: /metadata/idinfo/spdom/bounding/northbc/
Rule: not null
Domain: number between (90.00) and (-90.00)
- South Bounding Coordinate
Tag: /metadata/idinfo/spdom/bounding/southbc/
Rule: not null
Domain: number between (90.00) and (-90.00)
- Theme Keyword
Tag: /metadata/idinfo/keywords/theme/themekey/
Rule: Not null
Domain: free text
- Metadata Contact Organization
Tag: /metadata/metainfo/metc/cntinfo/cntorgp/cntorg/
Rule: not null if Metadata Contact Person is null
Domain: free text
- Metadata Contact Person
Tag: /metadata/metainfo/metc/cntinfo/cntperp/cntper/
Rule: not null if Metadata Contact Organization is null
Domain: free text

- Metadata Contact Address City
 Tag: /metadata/metainfo/metac/ctinfo/ctaddr/city/
 Rule: not null
 Domain: free text
- Metadata Contact Address State or Province
 Tag: /metadata/metainfo/metac/ctinfo/ctaddr/state/
 Rule: not null
 Domain: free text
- Metadata Contact Address Postal Code
 Tag: /metadata/metainfo/metac/ctinfo/ctaddr/postal/
 Rule: not null
 Domain: free text

Insertions

- Publication Date
 Tag: /metadata/idinfo/citation/citeinfo/pubdate/
 Rule: if null, insert 'unknown'
 Domain: "unknown", "unpublished material" or free date
 Date Format: YYYYMMDD (YYYY minimum)
- Purpose
 Tag: /metadata/idinfo/descript/purpose/
 Rule: if null, insert 'none provided'
 Domain: free text
- Time Period of Content: Single Date
 Tag: /metadata/idinfo/timeperd/timeinfo/sngdate/caldate
 Rule: if null and if Range of Dates and Multiple Dates are null,
 insert 'unknown'
 Domain: "unknown" or free date
 Date Format: YYYYMMDD (YYYY minimum)
- Time Period of Content: Range of Dates, Beginning Date
 Tag: /metadata/idinfo/timeperd/timeinfo/rngdates/begdate/
 Rule: if Ending Date is not null, insert 'unknown',
 Domain: "unknown" or free date
 Date Format: YYYYMMDD (YYYY minimum)
- Time Period of Content: Range of Dates, Ending Date
 Tag: /metadata/idinfo/timeperd/timeinfo/rngdates/enddate/
 Rule: if Beginning Date is not null, insert 'unknown',
 Domain: "unknown" or free date
 Date Format: YYYYMMDD (YYYY minimum)
- Currentness Reference
 Tag: /metadata/idinfo/timeperd/current/
 Rule: if null, insert 'unknown'
 Domain: free text
- Maintenance and Update Frequency
 Tag: /metadata/idinfo/status/update
 Rule: if null, insert 'unknown'
 Domain: free text

- Theme Keyword Thesaurus
Tag: /metadata/idinfo/keywords/theme/themekt/
Rule: if null, insert 'none'
Domain: free text
- Access Constraints
Tag: /metadata/idinfo/acconst/
Rule: if null, insert 'unknown'
Domain: free text
- Use Constraints
Tag: /metadata/idinfo/useconst/
Rule: if null, insert 'unknown'
Domain: free text
- Metadata Contact Address Type
Tag: /metadata/metainfo/metc/ctinfo/ctaddr/addrtype/
Rule: if null, insert 'unknown'
Domain: free text
- Metadata Contact Phone number
Tag: /metadata/metainfo/metc/ctinfo/ctvoice/
Rule: if null, insert 'unknown'
Domain: free text
- Metadata Date
Tag: /metadata/metainfo/metd
Rule: if null, insert harvest date
Domain: free date
Date Format: YYYYMMDD (YYYY minimum)

Appendix C

Sample XML Metadata Record with FGDC Essential Elements

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <metadata>
- <idinfo>
  - <citation>
    - <citeinfo>
      <origin>Louisiana State University Coastal Studies Institute</origin>
      <pubdate>20010907</pubdate>
      <title>Geomorphology and Processes of Land Loss in Coastal Louisiana, 1932 –
        1990</title>
    </citeinfo>
  </citation>
- <descript>
  <abstract>A raster GIS file that identifies the land loss process and
    geomorphology associated with each 12.5 meter pixel of land loss between
    1932 and 1990. Land loss processes are organized into a hierarchical
    classification system that includes subclasses for erosion, submergence,
    direct removal, and undetermined. Land loss geomorphology is organized
    into a hierarchical classification system that includes subclasses for both
    shoreline and interior loss.</abstract>
  <purpose>The objective of the study was to determine the land loss
    geomorphologies associated with specific processes of land loss in coastal
    Louisiana.</purpose>
</descript>
- <timeperd>
  - <timeinfo>
    - <rngdates>
      <begdate>1932</begdate>
      <enddate>1990</enddate>
    </rngdates>
    </timeinfo>
    <current>ground condition</current>
  </timeperd>
- <status>
  <progress>Complete</progress>
  <update>None planned</update>
</status>
- <spdom>
  - <bounding>
    <westbc>-92.000057</westbc>
    <eastbc>-88.81416</eastbc>
    <northbc>30.498417</northbc>
    <southbc>28.914905</southbc>
  </bounding>
</spdom>

```

```

- <keywords>
- <theme >
  <themekt >ISO 19115 Topic Category </themekt >
  <themekey>biota</themekey>
</theme >
- <theme >
  <themekt >none </themekt >
  <themekey>land loss</themekey>
  <themekey>wetlands</themekey>
  <themekey>geomorphology </themekey>
  <themekey>landscape ecology </themekey>
</theme >
</keywords>
<accconst >none </accconst >
<useconst >The metadata should be read completely prior to use of the data set.
  Data were collected and compiled as 12.5 meter pixels and should not be
  extended beyond the reasonable limits of the resolution. This is not a survey
  data product and should not be utilized as such. </useconst >
</idinfo>
- <metainfo>
  <metd>20010907 </metd>
- <metc>
  - <cntinfo>
    - <cntorgp>
      <cntorg>Louisiana State University Coastal Studies Institute</cntorg>
    </cntorgp>
    - <cntaddr>
      <addrtype>mailing and physical address</addrtype>
      <city>Baton Rouge</city>
      <state>LA</state>
      <postal>70803</postal>
    </cntaddr>
      <cntvoice>(225) 578-2395</cntvoice>
    </cntinfo>
  </metc>
  <metstdn>FGDC Content Standards for Digital Geospatial Metadata</metstdn>
  <metstdv>FGDC-STD-001-1998</metstdv>
</metainfo>
</metadata>

```