



Incorporating State Clearinghouse Nodes

FGDC Coordination Group



National States Geographic Information Council

Bill Burgess, Washington Liaison



Relevant to the Discussion

- The Content Standard for Digital Geospatial Metadata (CSDGM) was introduced in 1994
- We currently don't have a national system to manage everyone's metadata – we're 'floundering'
- A very large number of GIS Users across the nation (possibly 75%) simply don't document their data to the CSDGM Standard (or any other standard)
- The GIS Inventory was designed to help capture those users



Purpose of the GIS Inventory

- **Track the status of GIS in state and local government**
 - Provides the status of data development for planning purposes
 - Characterizes the User Community (many system & policy questions)
 - Can't manage what you don't understand
- **Locate data and users**
- Works in concert with federal programs for broad data discovery
- Provides a single national inventory tool
 - Reduces the need for the multiple inventories conducted by federal and state agencies



Not a Metadata tool

- Ramona is a management info system
- Most metadata systems produce card catalogs (not dependable for finding all information)
- We produce CSDGM compliant metadata because it is the standard for connecting to other systems
- We plan to import existing metadata from state clearinghouses and other tools, but this information will not fully populate the GIS Inventory database – DHS is funding this effort





Example from Wisconsin report based on the GIS Inventory

Orthoimagery

Foundational Element: Geographic Reference Frameworks

Framework Data Category: Orthoimagery

Statistics:

- 100% county-based coverage
- 4-5 year iterative planning cycle for many counties
- > 15 counties indicated “planned” imagery in 2010 – the actual number is suspected to be 2-4 times that number.
- Oblique imagery is rising in popularity, as well as in current and future investment.

Related Information:

With little ambiguity, it is clear that 100% of Wisconsin counties have invested in locally-funded orthoimagery sometime over the last 15 years – in some cases, in 4 or 5-year iterations. This aerial imagery rectified to ground control for integration in GIS systems has great value in collection, registration, and quality assurance of other foundational GIS layers.

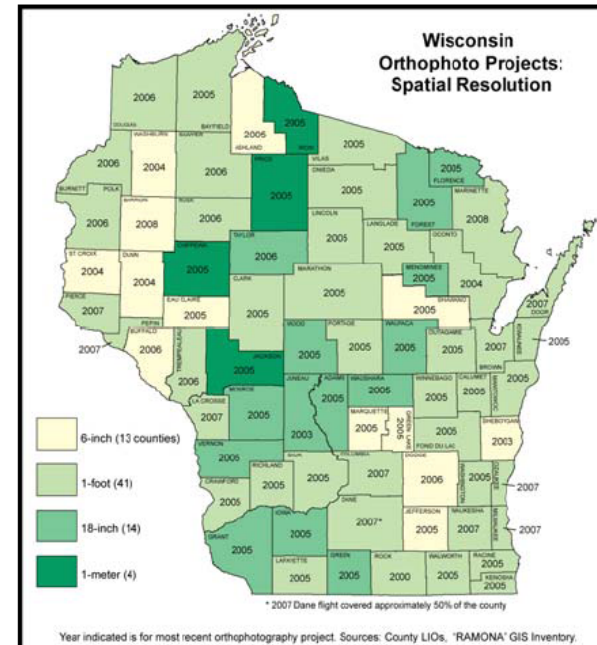


Figure 16 – Orthophoto Spatial Resolution

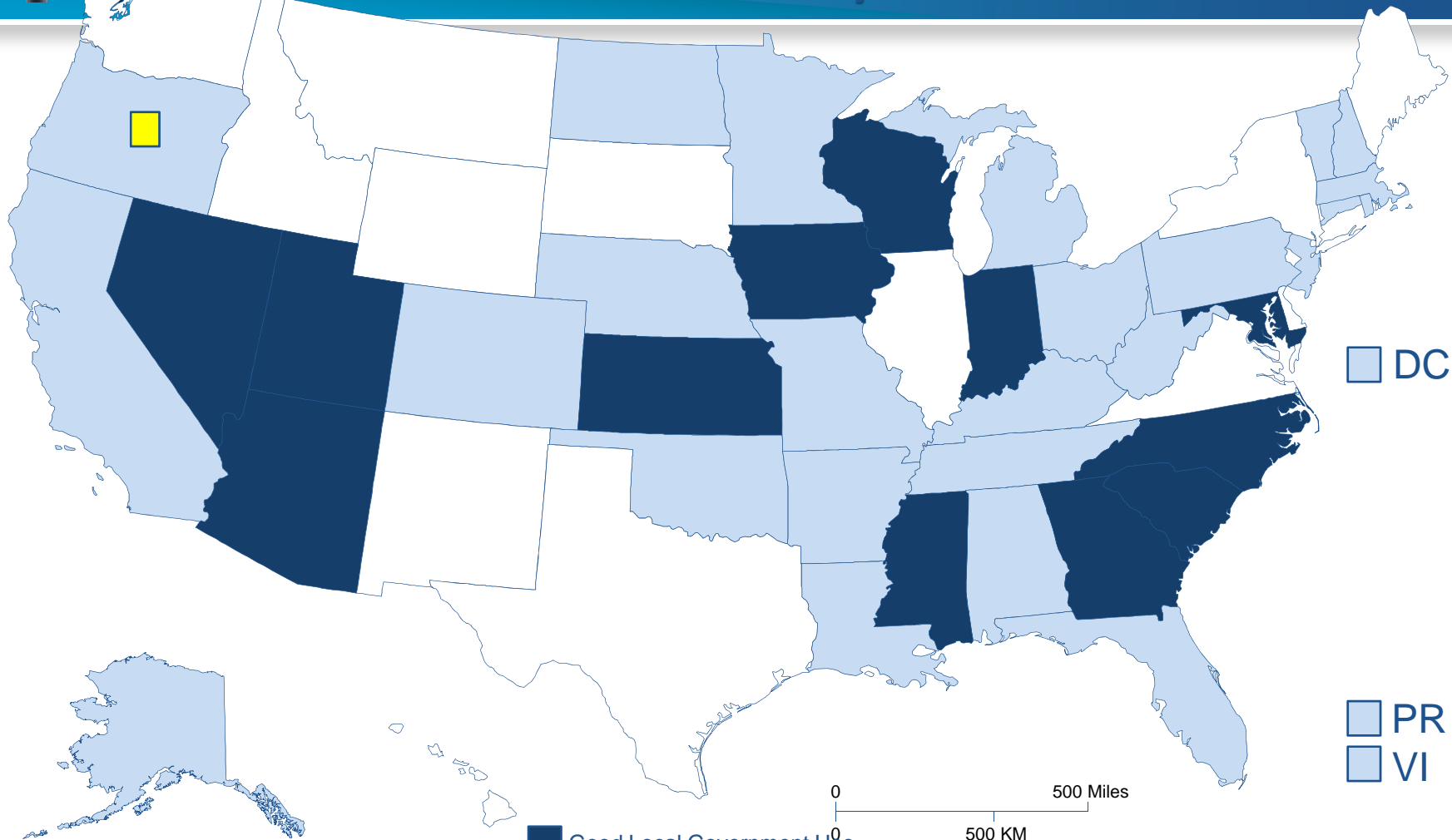


State Clearinghouse Nodes

- **Information from May 2011 provided by previous GOS Portal Team**
 - 60 Registered State Clearinghouse Nodes
 - 67 Harvested Sites
 - 59,385 Metadata Records
 - 18 have less than 25 metadata records
 - Idaho had 17,753 metadata records (30%)
 - Last harvest dates for nodes ranged from October 2005 to May 2011

Current Picture for Inventory

Subjective Measures



AK & HI not to scale





Issue 1: Name Game

Controlled Data Fields versus Free Text

Digital Orthophotography/Orthoimagery

- Digital Imagery
- Digital Photography
- Imagery
- Photography
- Aerial Photography
- Aerial Photos
- Aerial Imagery
- Satellite Imagery
- Orthoimagery
- Orthophotography
- Digital Orthoimagery
- Digital Orthophotography
- Quarter Quads
- DOQQs
- Digital Quads
- NAIP Imagery
- 133 Cities Imagery





Issue #2: Data Integration

- Controlled Fields
 - For data, we control entries for source material, scale, update frequency and other pieces of information
 - GIS Inventory may compromise on select fields to simplify ingestion
 - Remaining fields will very likely require one-time human interaction to import data from Clearinghouse sites



Issue #3: Scope of Metadata

• GIS Inventory

- Sections 1 & 7 Complete
- Partial information from Sections 2 through 6 is in the title and abstract

• Full CSDGM Metadata

- All 7 Sections

Ramona GIS Inventory Starter Metadata	
Title	100-Year Floodplains, This dataset is a polygon feature class showing 100 year flood data for Baltimore County, Published in 2008, 1:2400 (1in=200ft) scale, Baltimore County Government.
Origin(origin)	Baltimore County Government
Publication Date(pubdate)	2008
Publication Info(pubinfo)	Towson, MD Baltimore County Government
Other Citation(otherscit)	
Online Link(online)	http://www.baltimorecountymd.gov/Agencies/mnneighborhood/index.html
Online Link(online)	http://www.baltimorecountymd.gov/
Description	
Abstract(abstract)	This 100-Year Floodplains dataset, published at 1:2400 (1in=200ft) scale, was produced all or in part from Other information as of 2008. It is described as 'This dataset is a polygon feature class showing 100 year flood data for Baltimore County.'. Data by this publisher are often provided in Array coordinate system, in a Array projection; The extent of these data is generally Baltimore County, MD,Chesapeake Bay - Maryland, MD. This starter metadata was auto-generated through the Ramona GIS Inventory System (www.gisinventory.net) and does not represent all information about this data set.
Purpose(purpose)	This is an inventory-level metadata record documented through the GIS Inventory system. Please refer to the data set contact or publisher for original intent and appropriate use information.
Time Period	
Calendar Date(calendar)	2008
Currentness Reference(current)	Publication Date
Status	
Progress(progress)	Complete
Update(update)	Do Not Know

Title

100-Year Floodplains, This dataset is a polygon feature class showing 100 year flood data for Baltimore County, Published in 2008, 1:2400 (1in=200ft) scale, Baltimore County Government.

Abstract(abstract)

This 100-Year Floodplains dataset, published at 1:2400 (1in=200ft) scale, was produced all or in part from Other information as of 2008. It is described as 'This dataset is a polygon feature class showing 100 year flood data for Baltimore County.'. Data by this publisher are often provided in Array coordinate system; in a Array projection; The extent of these data is generally Baltimore County, MD,Chesapeake Bay - Maryland, MD. This starter metadata was auto-generated through the Ramona GIS Inventory System (www.gisinventory.net) and does not represent all information about this data set.



Issue #4: Scope of All Data

- **GIS Inventory**

- Managed as a database
- Focus is wide
 - User
 - Organization
 - Systems
 - Policies
 - Geography
 - Data

- **Clearinghouse Nodes**

- Managed as a catalog
- Focus is narrow
 - Data





Issue #5: Apples and Oranges

The screenshot shows the GIS Inventory website interface. The main map displays the United States with various colored regions. The legend on the right is titled 'Places That Produce' and includes the following items:

- Registered Users with Data of Any Type (Green circle)
- GIS Data (Grey circle)
- National Cadastral Data Infrastructure (Grey circle)

Below this, there is a 'Legend' section with color-coded boxes for different geographic levels:

- City/Town (Green box)
- County (Blue box)
- Statewide (Dark Blue box)
- Tribal (Red box)
- Coastal Areas (Yellow box)
- No Response (Grey box)



**GIS Inventory (GIS Data)
National Cadastral Data**

Operate Differently

**The same may be true for
the Clearinghouse Nodes**





Completed a Crosswalk

CSDGM GIS ISO Crosswalk 20120307 - Microsoft Excel

GIS Inventory Crosswalk to CSDGM and ISO 19115					
		*note field is from <i>My Data Layers</i> unless otherwise noted, e.g. <i>MyProfile:Organization</i>			vers. 20120302
GISI Metadata Record Elements	Content Source from GISI Data Entry Forms	CSDGM Field	CSDGM Domain	ISO Name	
Publication Info	Auto-concatonated from: <i>User City and Organization Name</i>	Publisher and Publication Place	free text	CI_Citation, gmd:citeResponsibleParty, gmd:CI_ResponsibleParty and CI_Citation, gmd:citeResponsibleParty, gmd:CI_ResponsibleParty, gmd:contactInfo, gmd:CI_Contact, gmd:address, gmd:CI_Address	
Other Citation	? What value populates this field?	Other Citation Details	free text	CI_Citation, gmd:otherCitationDetails	
Online Linkage (repeatable)	<i>Web Map URL and/or Full Metadata URL</i>	online linkage	free text	MD_Metadata.identificationInfo > MD_Identification.citation > CI_Citation.onlineResource > CI_OnlineResource	
Abstract	Auto-concatonated from: <i>Data Category, Approximate Scale, Source, Production Date, Description, SystemProfile:Map Coordinate System, SystemProfile:Map Projection, MyGeography:(fixed domain)</i>	Abstract	free text	MD_Metadata, identificationInfo, abstract	
Purpose	Fixed content: 'This is an inventory-level metadata record documented through the GIS Inventory System. Please refer to the data set contact or publisher for original intent and appropriate use information'	Purpose	free text	MD_Metadata, identificationInfo, purpose	



Completed Options – Now Costs



NSGIC GIS Inventory Options for Incorporating State, Tribal, and Local Metadata

Draft Report

Version 0.3
February 20, 2012

Met with the states to get their input and then circulated document for final comments.

NSGIC GIS Inventory - Options for State Metadata

This capability could be implemented as a synchronous or asynchronous process, with respect to the user interaction with the GIS system.

- Synchronous**—the system would immediately step the user through the process as outlined in Section 6.1.1, above.
- Asynchronous**—the system would parse the files, validate the files, populate a data holding record for each file, then notify the user (e.g., via email) that there are one or more pending steps to be performed for a number of records. The user could then come back to the system to address the remaining steps in the process, at their convenience.

6.1.3. Harvest from User CSW

Connection to an existing catalog service provided by user, harvested by GIS.

- States, Tribes, or localities provide a service endpoint to their CSW.
- GIS harvests with some periodicity.
- Associate the set of files in the harvestable directory with a particular user, for reasons stated above.
- Individual data holding records are created for each individual file that is uploaded.
- Asynchronous process would be followed, per description above in Section 6.1.2, 4b.

Specific potential technical solutions for this option are presented in Appendix 3 of this document.

6.1.4. Users Push to GIS Portal

GIS implements a portal that allows states, tribes, and localities to publish their metadata records directly to the GIS portal, "pushed by users".

- User pushes one or more metadata files to GIS portal, as new or updated files are available.
- GIS follows steps 3–5 listed in Section 6.1.3 (above).

Specific potential technical solutions for this option are presented in Appendix 3 of this document.

6.1.5. GIS Scrapes from Users

To provide greater flexibility, restrictions on users could be eased to allow them to make accessible metadata that is not strictly compliant to a standard, and not tightly structured.

- GIS can be extended to extract metadata suitable for inclusion into the inventory from sources such as relational database systems (RDBS), web accessible folders (WAF), and unstructured sources such as web pages and emails. Users can submit the source of metadata (along with any credentials required for access) to GIS which will then attempt to connect to the source and extract metadata.
- For RDBS extract, GIS will examine the database table structure(s) to determine what fields are populated. This requires users to submit the name of a "master" table which may contain foreign key references to other tables.

NSGIC GIS Inventory - Options for State Metadata

- For WAF extract, GIS will examine the folder(s) on disk to determine which files are suitable for extraction and then begin processing each file in a manner similar to web page/email extract (see below).
- For web page/email extract, GIS will access a URL (web page) or unstructured text document (email) and begin extraction using GIS's web crawler. Well-formed FGDC or ISO compliant metadata, existing data extraction techniques will be leveraged. If the content is not well formed, a unit-of-based approach will be applied to discern metadata values and the fields to which they pertain.
- GIS follows steps 3–5 listed in Section 6.1.3 (above).

Note that it may be necessary to break this enhancement option into 2 or more pieces, or otherwise use a phased approach, for example harvesting from existing WAFs in a "Phase 1".

6.1.6. API to GIS database

GIS would provide an API to the GIS database so that authorized users could more efficiently maintain their records in a more powerful and flexible manner (e.g. issue changes to all records given a change in an organization's name or the contact's name). This functionality defined here includes a capability of inserting new metadata records.

6.2. Secondary (or Other) Enhancement Options

A. Published "GIS profiles"

NSGIC defines and publishes CSDGM and ISO 19115 "GIS profiles". A GIS profile would define the profiling (sub-setting) of, and extensions (additions) to the base content standard. This would include a description of the extra fields (e.g. scope), constraints (fields (Name, Keywords) and fields generated by concatenation of other fields (Title and Abstract). If a possible file extension is defined for GIS-specific fields, e.g. GIS Title and GIS Abstract, so as to prevent discrepancies with elements in user's original metadata files. If GIS continues to handle only "light" metadata, then these profiles would become a considerable amount of sub-setting. If GIS decides to provide an administration of full metadata records, then the profiles would describe the extensions, constraints, and modified fields.

B. Full Content Integration

After importing an existing (original) metadata record into GIS (through one of the primary Enhancement options listed above), the full record would be stored in the system. This capability is fundamentally different than the generation of "light" metadata, with a URL link back to original record.

C. Metadata Standards Support

If the GIS only handled full metadata, and the CSDGM and ISO 19115 "GIS profiles" were declared ("A" above), then on those two standards would need to be supported. However, if GIS has already completed metadata records submitted into the system, then GIS would need to support the entire CSDGM, ISO 19115, and NAP content standards.

D. Metadata Format Transformation

Metadata could be presented in different formats through a template database view, or by applying transformations (e.g. XSL) to the metadata files (e.g. in XML format). This could be between the GIS profiles for only the "light" metadata, or it could be among those profiles as well as the entire CSDGM, ISO 19115, and NAP content standards if full content was being stored (not linked).



Next Steps

- Build basic function to ingest and manage State Clearinghouse information by September
 - Ability to export this information in CSDGM and ISO formats
- Include an incremental approach that will allow us to build additional functions into the System as funding is available



Questions ?

GIS INVENTORY

