



# Introduction to Metadata

Viv Hutchison  
US Geological Survey

April 20, 2011  
Denver, CO  
(via webex)



# Presenter

- Viv Hutchison
  - US Geological Survey
  - Core Science Systems - Biological Informatics Program
  - Metadata Program Coordinator
  - Location: USGS Center for Biological Informatics – Denver, CO
  - Contact: [vhutchison@usgs.gov](mailto:vhutchison@usgs.gov) or 303.202.4227



# Biological Informatics Program

**nbii** National Biological Information Infrastructure

Google™ Custom Search search

NBII Home About NBII Plants, Animals & Other Organisms Habitats Ecological Topics Geographic Perspectives Toolkit

## Early Detection, Rapid Response (EDRR)

A catalog of online invasive species resources, for identification, reporting, expertise, occurrences, assessment, planning, and response

PROGRAM AREAS 1 2 3 4 5

NBII: RSS Sources

- NBII Wildlife Disease Information Node's RSS Feeds
- NBII Mid-Atlantic Information Node's headlines RSS Feed
- NBII Pacific Basin Information Node's Environmental News Service RSS Feed
- NBII Southern Appalachian Information Node's RSS Feeds
- NBII Bird Conservation Node's RSS Feed
- Butterflies and Moths of North America (NBII-supported) blog on what's new in BAMONA with RSS

National Biological Information Infrastructure at a Glance

News from the NBII Blog

# Biological Informatics Program

## Explore NBII's Information Regionally:

NBII examines the priority natural resources challenges specific to various regions of the United States, and places special emphasis on providing content relevant to those challenges. Click a region to see a sample of the projects in that area.



# Overall Topics

- ❑ Metadata: What it is, Why it is valuable
- ❑ Challenges in Eco/Bio Informatics
- ❑ FGDC Content Standard for Digital Geospatial Metadata (CSDGM)
- ❑ Biological Data Profile
- ❑ Clearinghouses
- ❑ How to Write Metadata
- ❑ Implementation
- ❑ Beyond Discovery Metadata

If you don't know  
where you are going...

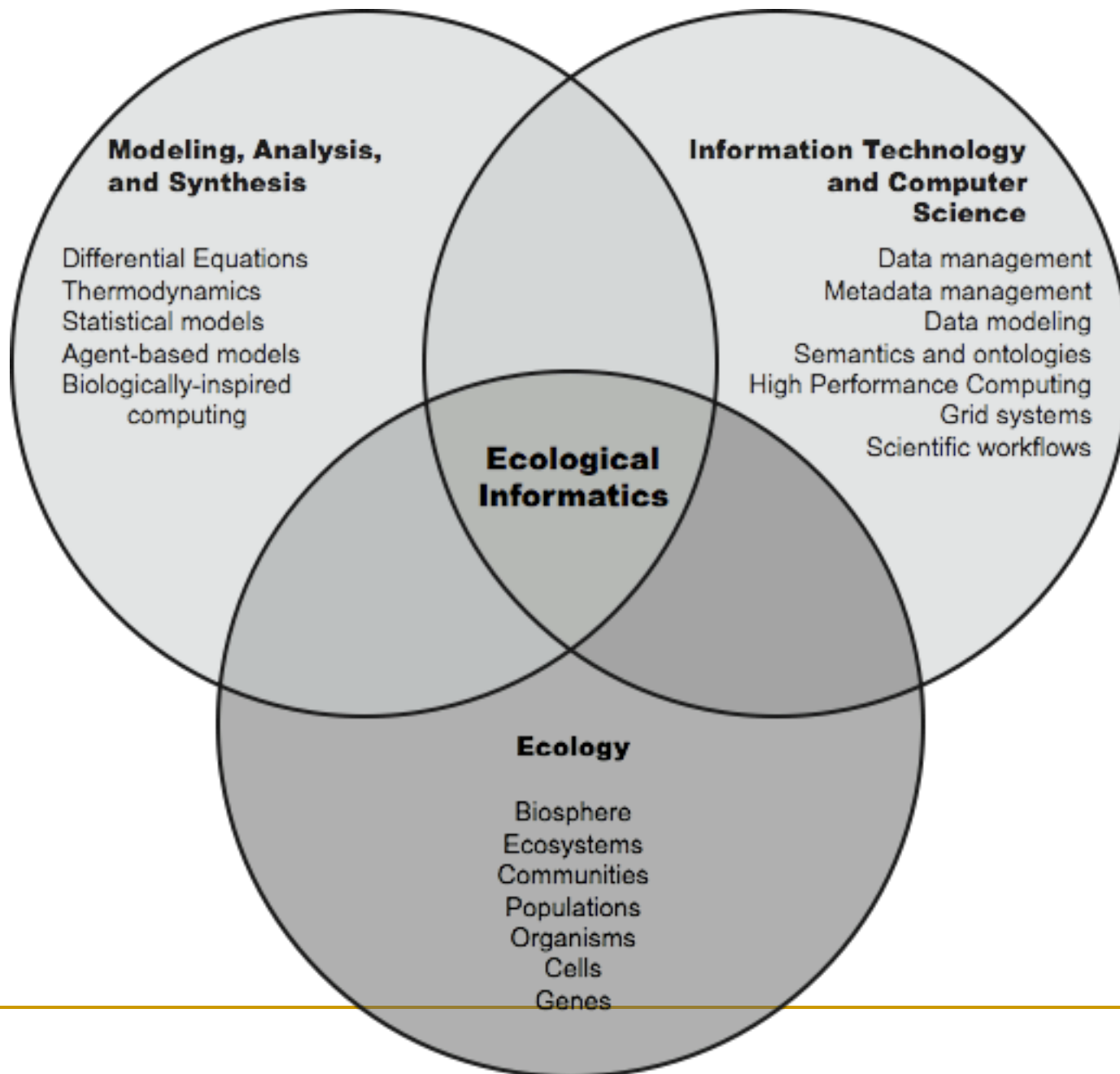
you'll wind up  
somewhere  
else.

—YOGI BERRA



---

# Bio/Eco Informatics and Challenges



---

# Informatics Challenges:

Majority of biological/ecological data is undocumented

- ❑ Lacks information on structure and content of data
- ❑ May be impossible to understand data without contacting the original researchers, which is problematic over the long-term

Documentation conventions widely vary

- ❑ Requires large time investment to understand each data set

Data loss

- ❑ Huge investments in research unavailable to future researchers and managers
-

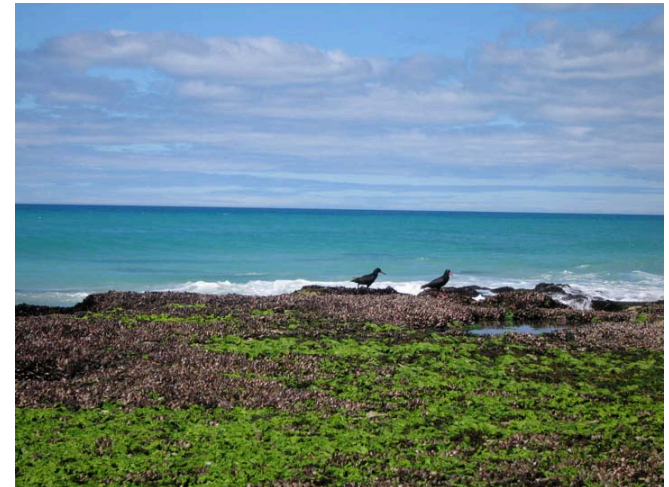
---

# Informatics Challenges, con't

- ❑ Scientists are not always well trained in the practice of data management
- ❑ Data management often not considered in academic curriculum (young professionals)
- ❑ Fading memory of completed datasets
- ❑ Hardware/software issues
- ❑ Archiving practices not considered
- ❑ Lost data or reduced data utility due to lack of data management practices
- ❑ Data are massively dispersed

# Data from multiple disciplines

- Wide variety of data used in Ecology and Biology
  - ❑ Population distribution surveys
  - ❑ Animal behavior data
  - ❑ Aerial photo surveys
  - ❑ Remote sensing data
  - ❑ Weather and Climate data
  - ❑ Hydrology data
  - ❑ Experimental data
  - ❑ Economics data
  - ❑ Human demographics
  - ❑ Land use and land cover data



DeHoop Nature Reserve, SA

# Data Integration Challenges

- Heterogeneity impedes integration

## Study A

METADATA (from EML)		<b>Study A:</b> White Mountains Area col. units: sq. meter PIRU = <i>Picea rubens</i> BEPA = <i>Betula papyifera</i>		
date	site	species	area	count
10/1/1993	N654	PIRU	2	26
10/3/1994	N654	PIRU	2	29
10/1/1993	N654	BEPA	1	3

## Study B

METADATA (from EML)		<b>Study B:</b> Green Mountains Area sampled: 1 sq. meter picrub = <i>Picea rubens</i> betpap = <i>Betula papyifera</i>	
date	site	picrub	betpap
31 Oct 1993	1	13.5	1.6
14 Nov 1994	1	8.4	1.8

## Integrated Data

study	date	site	species	density
A	10/1/1993	N654	<i>Picea Rubens</i>	13.0
A	10/3/1994	N654	<i>Picea Rubens</i>	14.5
A	10/1/1993	N654	<i>Betula papyifera</i>	3.0
B	10/31/1993	1	<i>Picea Rubens</i>	13.5
B	10/31/1993	1	<i>Betula papyifera</i>	1.6
B	11/14/1994	1	<i>Picea Rubens</i>	8.4
B	11/14/1994	1	<i>Betula papyifera</i>	1.8

metadata  
'promoted'  
to become  
data

format  
normalized  
using  
metadata

species metadata  
from study B  
is now data  
(picrub/betpap  
column headings)

density  
calculated  
using  
metadata

# Cultural Barriers to Data Sharing

- Research-based careers have reward system based upon individual scientific output
- Perceived risks of sharing data still exist
- “Sharing Publication-related Data and Material” published by the National Research Council, 2003:
  - Four Principles:
    - Authors should include data with their publications
    - If it can't be included, should be made freely available
    - Data should be placed in public repositories
    - Authors should anticipate which data/materials will be requested and state where they can be obtained

---

# Data Management

- What is data management?
  - Data management is the planning, execution of policies, practices, and projects that acquire, control, protect, deliver, and enhance the value of data and information assets.
- DAMA International, *The DAMA Guide to the Data Management Body of Knowledge*

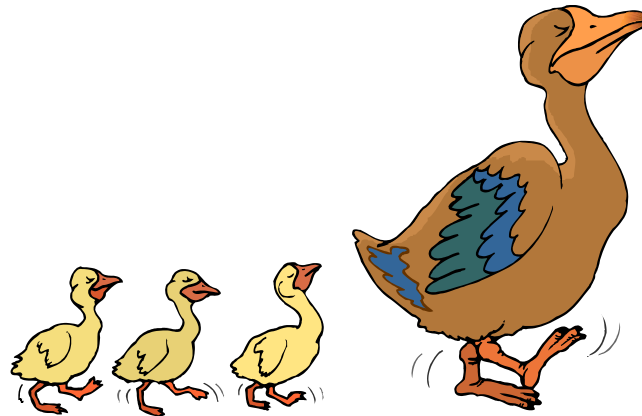
# Why Data Management?

- Why data management?
  - To understand the information needs of the enterprise and its stakeholders
  - To capture, store, protect, and ensure the integrity of data assets
  - To continually improve the quality of data and information including: data accuracy, integrity, integration, timeliness of data capture and presentation, relevance and usefulness of data
  - Ensure appropriate use of data and information
  - Maximize the effective use and value of data and information assets

- DAMA International, *The DAMA Guide to the Data Management Body of Knowledge*

---

# Getting Started with Metadata...

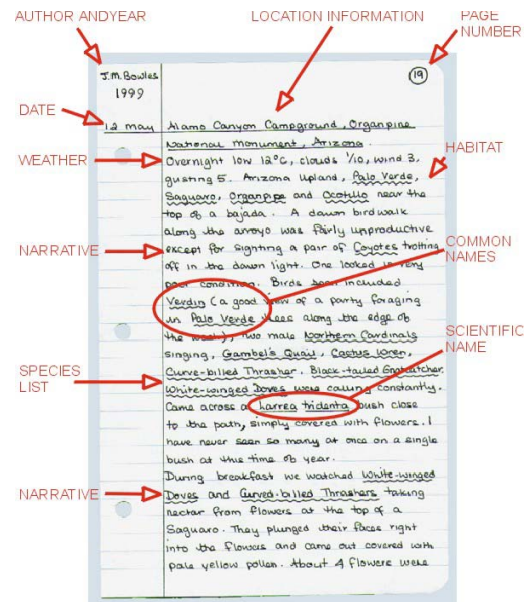
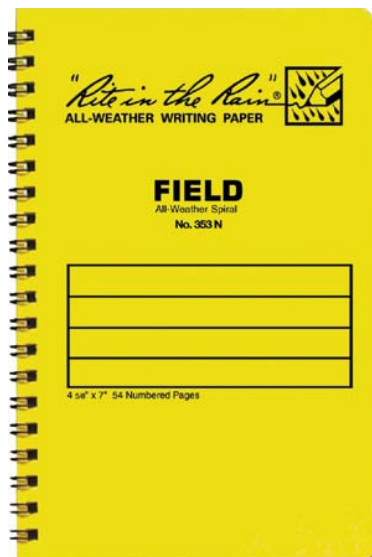
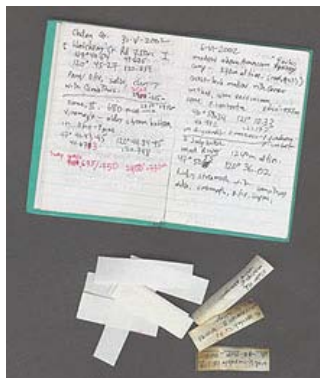


---

# Topics

- Define metadata
- Look at examples of metadata
- Determine the types of information included in a metadata record

# Data Collection



Birds found near Carlisle Pa		Birds found at Carlisle Pa		First Met	Years Met
Common Name	When first seen	Common Name	When first seen		
Lark Pigeon		Warblers Yellow breasted		5 <sup>th</sup> July 1839	13 yrs
Butcher Turkey	26 <sup>th</sup> Decr 1839	Wren eating		28 <sup>th</sup> July 1839	13 yrs
Coast White-headed	"	Yellow Redpoll	xxx "	27 <sup>th</sup> July 1839	13 yrs
Coast Red Tailed	8 <sup>th</sup> Sept 1836	Chatterer Cedar Bird		Aug 1832	13 yrs
XX "	March	Thrushes Mocking Bird		Jan 7 1840	10 yrs
"	Sparrow	"	Terrestrial	July 1830	13 yrs
XXX "	Pigeon	"	Cat Pigeon	Aug 1832	10 yrs
XY "	Slate Coloured	"	American Robin	July 1830	13 yrs
"	Fish Hawk	"	Wilson	23 <sup>rd</sup> April 1839	13 yrs
XX "	Winter	"	10 <sup>th</sup> mit	18 <sup>th</sup> July 1839	13 yrs
Onl "	Great Horned	"	Blain Ormou	4 <sup>th</sup> July 1839	13 yrs
"	Barred	"	Mate	27 <sup>th</sup> April 1839	13 yrs
"	Mottled	"	Wren	4 <sup>th</sup> May 1839	13 yrs
"	"	Cuckoo Yellow billed		1 <sup>st</sup> July 1837	12 yrs
"	"	Woodpecker	White	2 <sup>nd</sup> Jan 1840	6 yrs

# From Field Notes to Data Sets

## Average Temperature of Observation for Each Species

Species	Average Temperature	Temperature Standard Deviation	Number of Observations	Minimum Temperature	Maximum Temperature
Northern Red-legged Frog	4.4	---	1	4.4	4.4
Tailed Frog	7.0	3.0	3	4	10
Arizona Toad	10.0	---	1	10	10
Strecker's Chorus Frog	10.5	2.0	11	9	16
Oregon Spotted Frog	11.0	15.5	2	0	22
New Jersey Chorus Frog	11.5	4.5	17	3	22
Wood Frog	12.5	5.5	897	0	28.8
Spring Peeper	13.2	5.6	569	-1	32
Red-legged Frog	13.3	5.9	16	4	27

---

# From Data Sets to Published Papers



---

# What is a Data Set?

- A collection of data that is raw or statistically analyzed
  - **Generally data sets are defined as:**
    - Spatial – a collection of logically related features arranged in a prescribed manner such as GIS map layers, water features, etc
    - Tabular – a file, spreadsheet, data in a table
  - **Elements in a data set can include:**
    - Values, measures, points, coordinates, conditions, qualities, frequencies, or attributes that are a result of an observational study

# What Is Metadata?

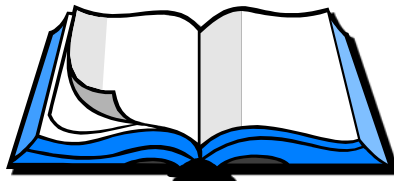
## ■ Metadata is: Data 'reporting'

- ❑ **WHO** created the data?
- ❑ **WHAT** is the content of the data?
- ❑ **WHEN** was it created?
- ❑ **WHERE** is it geographically?
- ❑ **HOW** was the data developed?
- ❑ **WHY** was the data developed?



# Metadata in the Real World

- *Metadata is all around...*



<b>Nutrition Facts</b>	
Serving Size ½ cup (114g)	
Servings Per Container 4	
<b>Amount Per Serving</b>	
<b>Calories</b> 90	Calories from Fat 30
<b>% Daily Value*</b>	
<b>Total Fat</b> 3g	<b>5%</b>
Saturated Fat 0g <b>0%</b>	
<b>Cholesterol</b> 0mg	<b>0%</b>
<b>Sodium</b> 300mg	<b>13%</b>
<b>Total Carbohydrate</b> 13g	<b>4%</b>
Dietary Fiber 3g <b>12%</b>	
Sugars 3g	
<b>Protein</b> 3g	
Vitamin A 80%	Vitamin C 60%
Calcium 4%	Iron 4%
* Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:	
	Calories: 2,000    2,500
Total Fat	Less than 65g    80g
Sat Fat	Less than 20g    25g
Cholesterol	Less than 300mg    300mg
Sodium	Less than 2,400mg    2,400mg
Total Carbohydrate	300g    375g
Dietary Fiber	25g    30g
Calories per gram:	
Fat 9 • Carbohydrate 4 • Protein 4	

<b>Author(s)</b>	Boullosa, Carmen.
<b>Title(s)</b>	They're cows, we're pigs / by Carmen Boullosa
<b>Place</b>	New York : Grove Press, 1997.
<b>Physical Descr</b>	viii, 180 p ; 22 cm.
<b>Subject(s)</b>	Pirates Caribbean Area Fiction.
<b>Format</b>	Fiction

# Working With Data

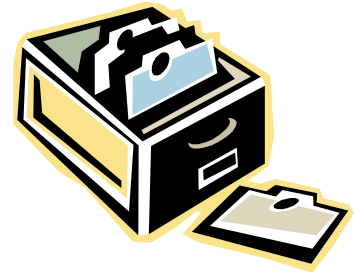


- When you **provide** data to someone else, what types of information would you want to include with the data?



- When you **receive** a dataset from an external source, what types of information do you want to know about the data?

# Working With Data



## ■ Providing data:

- ❑ Why was the data was created?
- ❑ What limitations, if any, do the data have?
- ❑ What does the data mean?
- ❑ Who should be cited if someone publishes something that utilized your data?

## ■ Receiving data:

- ❑ What are the data gaps?
- ❑ What processes were used for creating the current data?
- ❑ Are there any fees associated with the data?
- ❑ In what scale were the data created?
- ❑ What do the values in the tables mean?
- ❑ What software do I need in order to read the data?
- ❑ What projection is the data in?
- ❑ Can I give this data to someone else?

# What is a metadata standard?

*A Standard provides a structure to describe data with:*

- **Common terms to allow consistency between records**
- **Common definitions for easier interpretation**
- **Common language for ease of communication**
- **Common structure to quickly locate information**



*In search and retrieval, standards provide:*

- ▣ **documentation structure in a reliable and predictable format for computer interpretation**
- ▣ **a uniform summary description of the data set**

# What does a metadata record look like?

## North American Breeding Bird Survey (BBS)

### Identification Information:

#### Citation:

##### Citation Information:

*Originator:* Patuxent Wildlife Research Center, Biological Resources Division, U.S. Geological Survey (USGS)

*Publication Date:* 1997

##### Title:

North American Breeding Bird Survey (BBS)

##### Publication Information:

*Publication Place:* Laurel, MD

##### Publisher:

Patuxent Wildlife Research Center, Biological Resources Division, U.S. Geological Survey (USGS)

##### Other Citation Details:

This metadata file can be found at: [ftp://cameron.cr.usgs.gov/pub/nbii\\_metadata/brdpwrc0004.txt](ftp://cameron.cr.usgs.gov/pub/nbii_metadata/brdpwrc0004.txt) (text format) and

[ftp://cameron.cr.usgs.gov/pub/nbii\\_metadata/brdpwrc0004.html](ftp://cameron.cr.usgs.gov/pub/nbii_metadata/brdpwrc0004.html) (HTML format) and

[ftp://cameron.cr.usgs.gov/pub/nbii\\_metadata/brdpwrc0004.sgml](ftp://cameron.cr.usgs.gov/pub/nbii_metadata/brdpwrc0004.sgml) (SGML format).

### Description:

#### Abstract:

The North American Breeding Bird Survey (BBS), which is coordinated by the Biological Resources Division and Canadian Wildlife Service, is a primary source of population trend and distribution information for most species of North American birds. The BBS was initiated during 1966 by Chan Robbins and his associates at the Patuxent Wildlife Research Center to monitor the populations of all breeding bird species across the continental U.S., Canada, and Alaska. Approximately 2200 skilled observers participate in the survey each year. The BBS has accumulated 30 years of data on the abundance, distribution, and trends for more than 400 species of birds. These data are widely used by research agencies, non-governmental organizations, and the general public. Analyses of BBS data have been instrumental in the development of innovative approaches for analyzing trends of wild

#### Purpose:

In the 1960's, chlorinated hydrocarbon pesticides and similar poisons were widely used for spraying not only killed insects but also killed birds, raising serious concerns over its effects on birds. Unfortunately, no long-term regional or continental population data were available for birders to demonstrate declines in bird populations. The Bird Breeding Survey has provided information on bird population trends. Robbins et al. (1986) provided the first continental survey of songbirds based on BBS data. When viewed at continental or regional scales, these data provide information of the relative abundance of species that are well sampled by the BBS. In addition, they provide temporal patterns in trends. Populations of permanent resident and short-distance migrants

*Place Keyword:* Mexico

#### Taxonomy:

##### Keywords/Taxon:

*Taxonomic Keyword Thesaurus:* none

*Taxonomic Keyword:* birds

*Taxonomic Keyword:* breeding birds

##### Taxonomic Classification:

*Taxon Rank Name:* Kingdom

*Taxon Rank Name:* Animalia

##### Taxonomic Classification:

*Taxon Rank Name:* Phylum

*Taxon Rank Name:* Chordata

##### Taxonomic Classification:

*Taxon Rank Name:* Class

*Taxon Rank Name:* Aves

##### Taxonomic Classification:

*Taxon Rank Name:* Order

##### Taxonomic Classification:

*Taxon Rank Name:* Family

##### Taxonomic Classification:

*Taxon Rank Name:* Genus

##### Taxonomic Classification:

*Taxon Rank Name:* Species

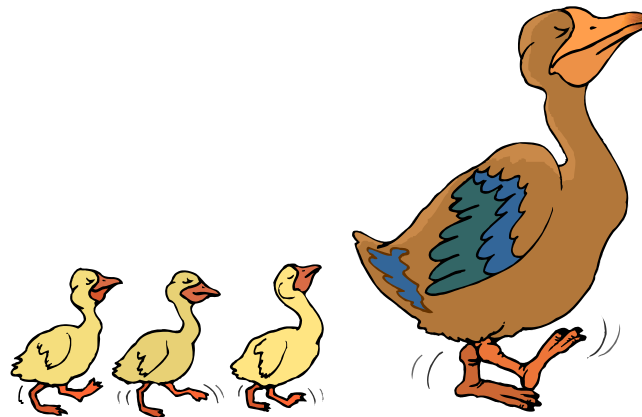
---

# Multiple metadata standards exist

- **Dublin Core Element Set**
  - Emphasis on web resources, publications
- **Content Standard for Digital Geospatial Metadata (CSDGM)**
  - Federal Geographic Data Committee (FGDC)
  - Emphasis on geospatial data
- **Biological Data Profile (BDP) of the CSDGM**
  - Emphasis on biological data (and geospatial)
- **ISO 19115 Geographic information: Metadata**
  - Emphasis on geospatial data and services
- **Darwin Core**
  - Emphasis on museum specimens
- **Ecological Metadata Language**
  - Emphasis on ecological data documentation

---

# The Value of Metadata: Why Do It?



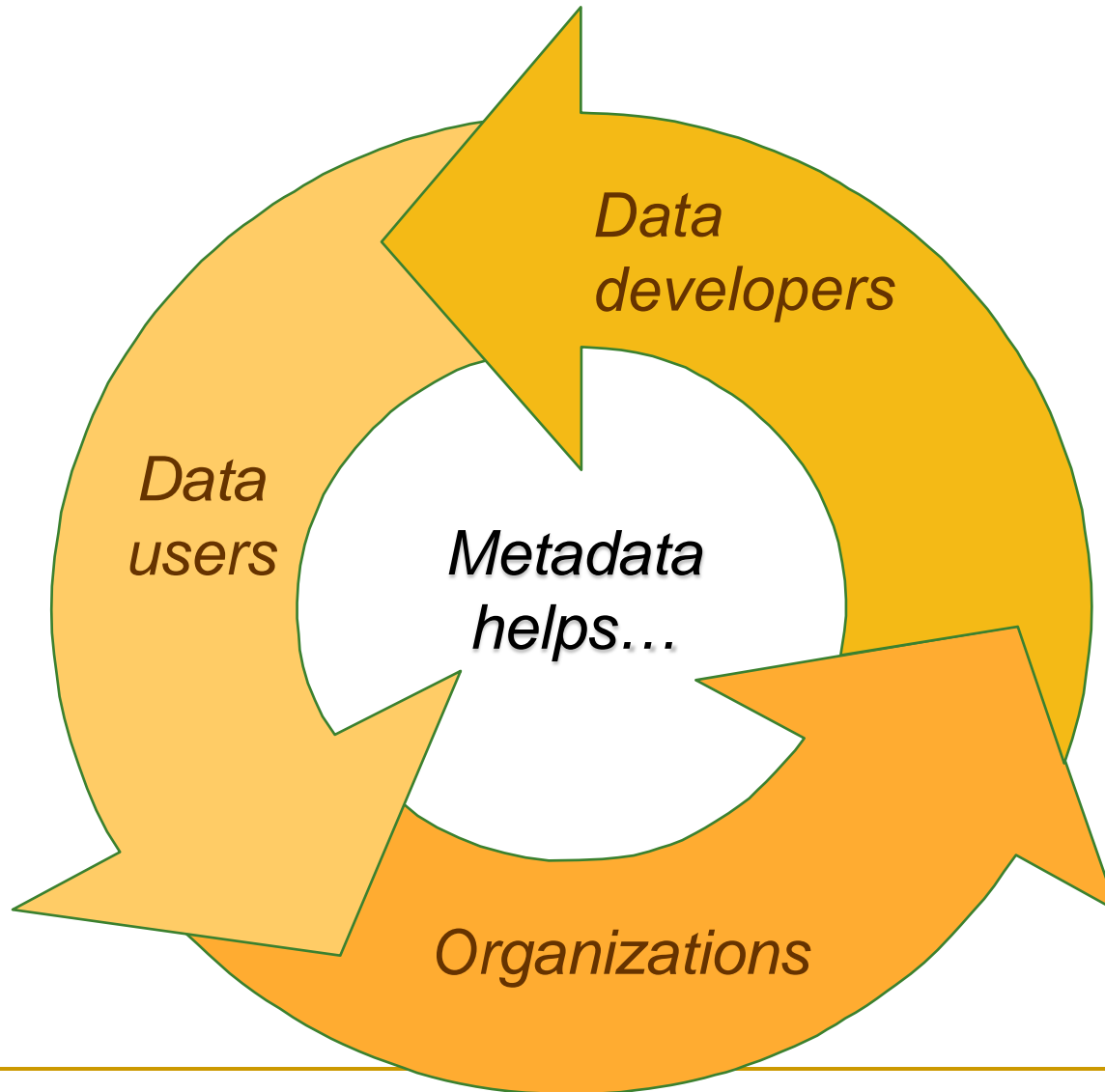
---

# Topics

- Determine the value of metadata
- Define reasons for a standard

---

Metadata has value to all:



# What is the value to *Data Developers*?

- Metadata allows data developers to:
  - ❑ Avoid data duplication
  - ❑ Share reliable information
  - ❑ Publicize efforts – promote scientist's work and contributions to a field of study
  - ❑ Reduce workload



# What is the value to *Data Users*?

- Metadata gives a user the ability to:
  - ❑ Search, retrieve, and evaluate dataset information from both inside and outside an organization
  - ❑ Find data - determine what data exist for a geographic location and/or topic
  - ❑ Determine applicability – decide if a data set meets your needs
  - ❑ Discover how to acquire the data set you identified, process and use the data set

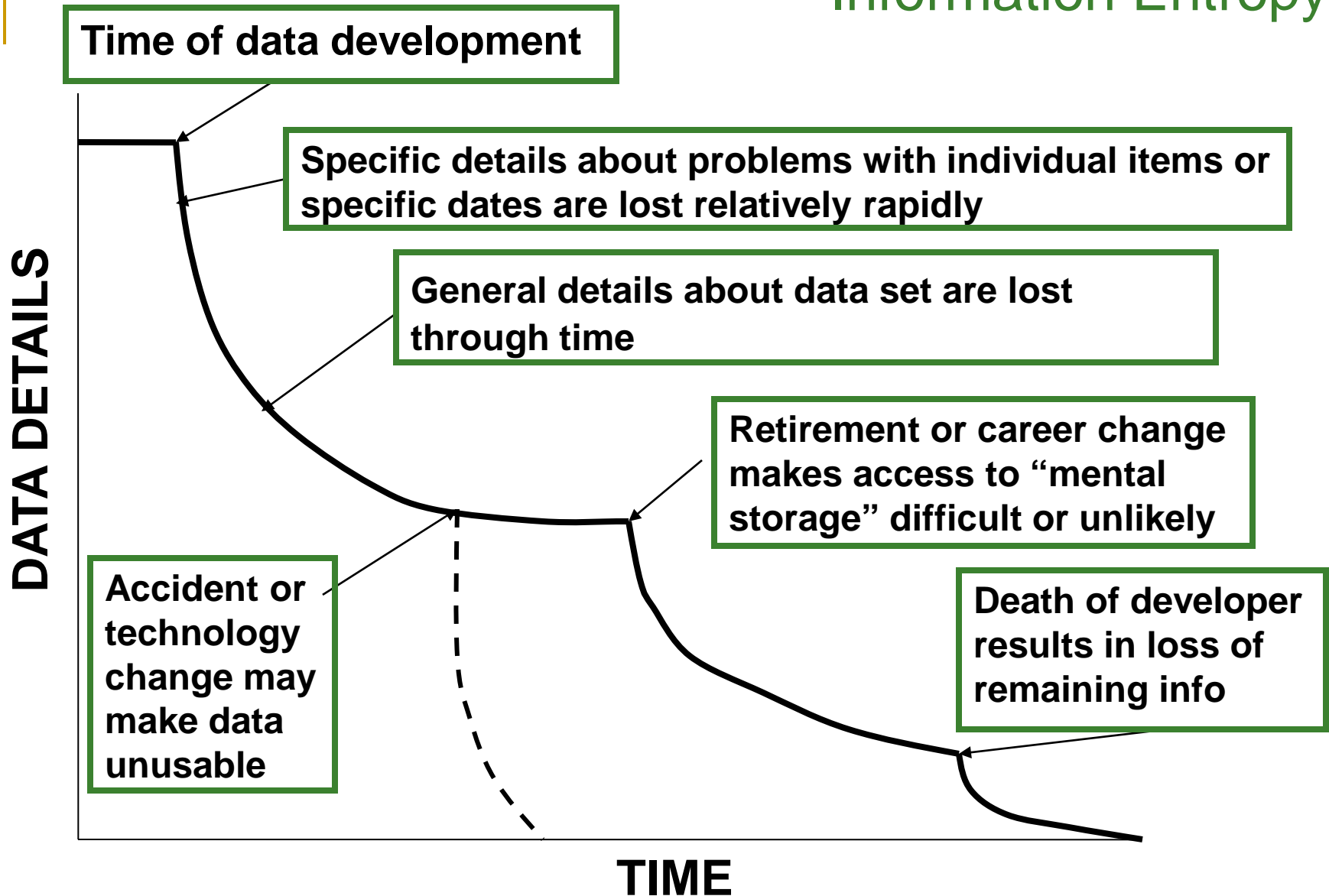


# What is the value to *Organizations*?

- Metadata helps ensure an organization's investment in data:
  - Documentation of data processing steps, quality control, definitions, data uses, and restrictions
  - Ability to use data after initial intended purpose
- Transcends people and time:
  - Offers data permanence
  - Creates institutional memory
- Advertises an organization's research
  - Creates possible new partnerships and collaborations thru data sharing



# Information Entropy



(From Michener et al 1997)

# Information Entropy

DATA DETAILS

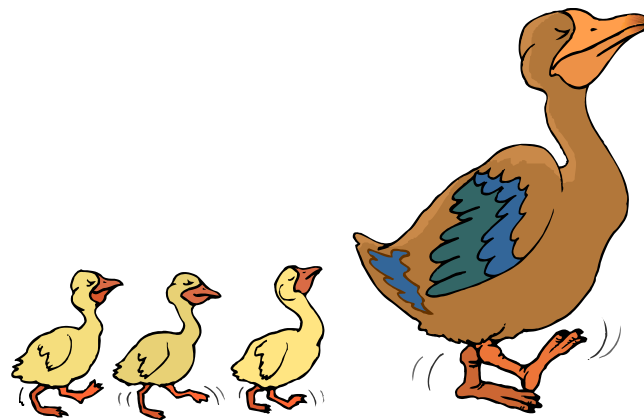
Sound information management –  
including metadata development  
can arrest the loss of data set detail



TIME

---

# A little history...



---

# Topics

- United States Executive Order 12906
- Metadata responsibilities for U.S. Federal agencies
- Federal Geographic Data Committee (FGDC)

# Executive Order 12906



## The Executive Order:

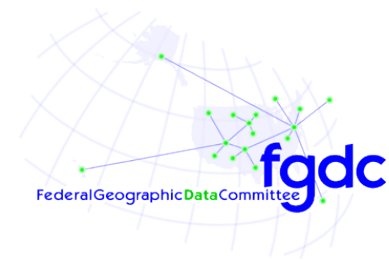
- Signed in 1994 by then U.S. President Clinton
- Defines the responsibilities of the Federal Geographic Data Committee (FGDC)
- Outlines three major uses of metadata:
  - (1) to maintain an organization's *internal investment* in geospatial data
  - (2) to provide information to data *clearinghouses* and catalogs
  - (3) to provide information needed to process and interpret data *transferred* from another organization.
- Requires creation of metadata for data sets from 1995 forward

# The Federal Geographic Data Committee

- The FGDC is a 19 member interagency committee working to implement the six basic building blocks of the National Spatial Data Infrastructure (NSDI): metadata, clearinghouse, standards, framework, geospatial data, and partnerships.

## Responsibilities:

- development of the NSDI
- establishment of the National Geospatial Data Clearinghouse
- cooperative efforts with state, local, and tribal governments, and the private sector
- implementation of digital geospatial data framework
- development of standards



you are here: [home](#)


## The Federal Geographic Data Committee

The Federal Geographic Data Committee (FGDC) is an interagency committee that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis. This nationwide data publishing effort is known as the [National Spatial Data Infrastructure](#) (NSDI). The NSDI is a physical, organizational, and virtual network designed to enable the development and sharing of this nation's digital geographic information resources. FGDC activities are administered through the FGDC Secretariat, hosted by the [National Geospatial Programs Office](#) (NGPO) of the U.S. Geological Survey. The NGPO oversees other geospatial programs of national importance including The National Map and the Geospatial One-Stop activity.


The Office of Management and Budget (OMB) established the FGDC in 1990 and rechartered the committee in its August 2002 revision of Circular A-16, "Coordination of Geographic Information and Related Spatial Data Activities." The FGDC is a 19 member interagency committee composed of representatives from the Executive Office of the President, and Cabinet level and independent Federal agencies. The Secretary of the Department of the Interior chairs the FGDC, with the Deputy Director for Management, Office of Management and Budget (OMB) as Vice-Chair. Numerous stakeholder organizations participate in FGDC activities representing the interests of state and local government, industry, and professional organizations.

- [What's new at the FGDC?](#)
- [Organization of the FGDC](#)

### News

 [NRC study report: Successful Response Starts with a Map](#)  
February 13, 2007

 [FGDC 2006 Winter Newsletter now available](#)  
February 07, 2007

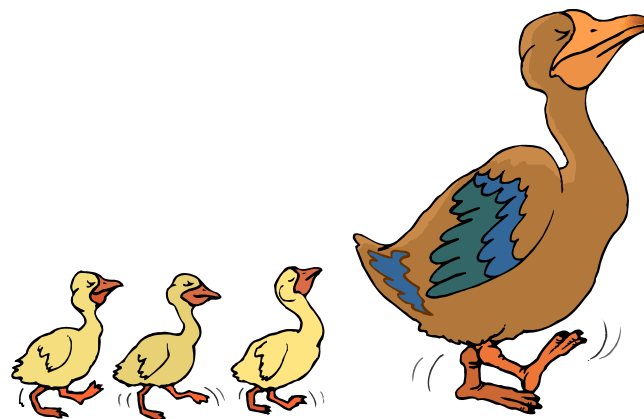
 [Extension of 2007 NSDI CAP-Category 4](#)  
February 05, 2007

 [2006 FGDC Annual Report Draft Review, Comments & Graphics due Feb 16, 2007](#)  
February 02, 2007

 [Metadata: North American Profile Committee Review](#)

---

# Standards, Profiles, Extensions, and ISO...



---

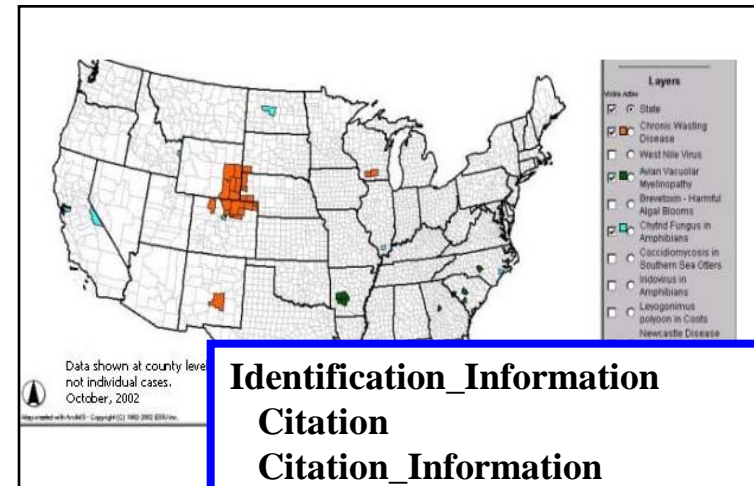
# Topics

- Distinguish between ‘FGDC Content Standard’ and ‘Profiles & Extensions’
- Introduce NBII Biological Data Profile
- Update: status of ISO standard

# FGDC Standard

## Content Standard for Digital Geospatial Metadata

The Content Standard for Digital Geospatial Metadata (CSDGM), Version 2 (FGDC-STD-001-1998) is the US Federal Metadata standard. The Federal Geographic Data Committee originally adopted the CSDGM in 1994 and revised it in 1998.



### Identification Information

Citation

Citation Information

Originator:

Publication Date

Title:

Geospatial Data Presentation Form

Publication Information

Publication Place

Publisher

Larger Work Citation

Citation Information

---

# FGDC metadata standard: Overview

## **Seven Major Metadata Sections:**

- Section 1 - Identification Information\*
- Section 2 - Data Quality Information
- Section 3 - Spatial Data Information
- Section 4 - Spatial Reference Information
- Section 5 - Entity and Attribute Information
- Section 6 - Distribution Information
- Section 7 - Metadata Information\*

## **Three Supporting Sections:**

- Section 8 - Citation Information\*
- Section 9 - Time Period Information\*
- Section 10 - Contact Information\*

***\* Minimum required metadata***

---

# FGDC Profiles and Extensions

**Profile:** a group of elements, added into the CSDGM standard in appropriate places, that are approved by the FGDC and address a specific community. A profile may also change the conditionality of an element (ie: mandatory, optional)

*Examples:* Biological Data Profile, Shoreline Data Profile

**Extension:** added elements outside the standard and needed by individual data set producers to customize the standard

*Examples:* Remote Sensing Extension

# Biological Data Profile

Additional elements:

- Taxonomy
- Methodology
- Analytical tools



# Biological Data Profile

Documents three types of datasets:

- Explicitly biological
- Biological and geospatial
- Explicitly geospatial



# The FGDC Standard with elements of the BDP

## North American Breeding Bird Survey (BBS)

### Identification Information:

#### Citation:

##### Citation Information:

*Originator:* Patuxent Wildlife Research Center, Biological Resources Division, U.S. Geological Survey (USGS)

*Publication Date:* 1997

##### Title:

North American Breeding Bird Survey (BBS)

##### Publication Information:

*Publication Place:* Laurel, MD

##### Publisher:

Patuxent Wildlife Research Center, Biological Resources Division, U.S. Geological Survey (USGS)

##### Other Citation Details:

This metadata file can be found at: [ftp://cameron.cr.usgs.gov/pub/nbii\\_metadata/brdpwrc0004.txt](ftp://cameron.cr.usgs.gov/pub/nbii_metadata/brdpwrc0004.txt) (text format) and

[ftp://cameron.cr.usgs.gov/pub/nbii\\_metadata/brdpwrc0004.html](ftp://cameron.cr.usgs.gov/pub/nbii_metadata/brdpwrc0004.html) (HTML format) and

[ftp://cameron.cr.usgs.gov/pub/nbii\\_metadata/brdpwrc0004.sgml](ftp://cameron.cr.usgs.gov/pub/nbii_metadata/brdpwrc0004.sgml) (SGML format).

### Description:

#### Abstract:

The North American Breeding Bird Survey (BBS), which is coordinated by the Biological Resources Division and Canadian Wildlife Service, is a primary source of population trend and distribution information for most species of North American birds. The BBS was initiated during 1966 by Chan Robbins and his associates at the Patuxent Wildlife Research Center to monitor the populations of all breeding bird species across the continental U.S., Canada, and Alaska. Approximately 2200 skilled observers participate in the survey each year. The BBS has accumulated 30 years of data on the abundance, distribution, and trends for more than 400 species of birds. These data are widely used by research agencies, non-governmental organizations, and the general public. Analyses of BBS data have been instrumental in the development of innovative approaches for analyzing trends of wild

#### Purpose:

In the 1960's, chlorinated hydrocarbon pesticides and similar poisons were widely used for spraying not only killed insects but also killed birds, raising serious concerns over its effects on birds. Unfortunately, no long-term regional or continental population data were available for birders to demonstrate declines in bird populations. The Bird Breeding Survey has provided information on bird population trends. Robbins et al. (1986) provided the first continental survey of songbirds based on BBS data. When viewed at continental or regional scales, these data provide information on the relative abundance of species that are well sampled by the BBS. In addition, they provide information on temporal patterns in trends. Populations of permanent resident and short-distance migratory

*Place Keyword:* Mexico

#### Taxonomy:

##### Keywords/Taxon:

*Taxonomic Keyword Thesaurus:* none

*Taxonomic Keyword:* birds

*Taxonomic Keyword:* breeding birds

##### Taxonomic Classification:

*Taxon Rank Name:* Kingdom

*Taxon Rank Name:* Animalia

##### Taxonomic Classification:

*Taxon Rank Name:* Phylum

*Taxon Rank Name:* Chordata

##### Taxonomic Classification:

*Taxon Rank Name:* Class

*Taxon Rank Name:* Aves

##### Taxonomic Classification:

*Taxon Rank Name:* Order

##### Taxonomic Classification:

*Taxon Rank Name:* Family

##### Taxonomic Classification:

*Taxon Rank Name:* Genus

##### Taxonomic Classification:

*Taxon Rank Name:* Species

# Biological Data Profile Workbook



National Spatial Data Infrastructure

## **Biological Data Profile Workbook**

(For use with FGDC-STD-001-1998)

**Federal Geographic Data Committee**

**March 30, 2001**

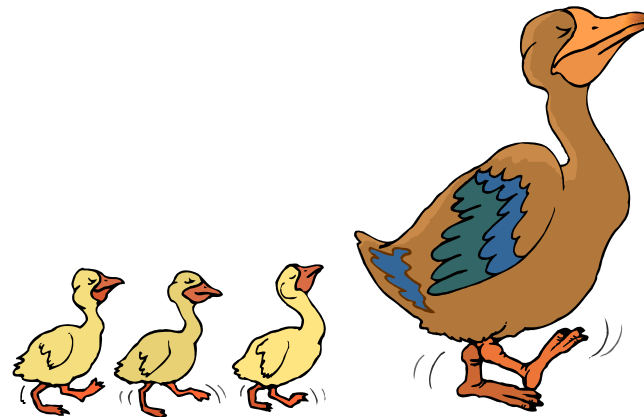
---

Federal Geographic Data Committee

Department of Agriculture • Department of Commerce • Department of Defense • Department of Energy  
Department of Health & Human Services • Department of Housing and Urban Development  
Department of the Interior • Department of Justice • Department of State  
Department of Transportation • Environmental Protection Agency  
Federal Emergency Management Agency • Library of Congress  
National Aeronautics and Space Administration • National Archives and Records Administration  
National Science Foundation • Tennessee Valley Authority

---

# FGDC to ISO: The Transition



# International Organization for Standardization (ISO) Metadata Standard

- The international community, through the International Organization for Standardization (ISO) designed a standard for geospatial metadata.
- Essentially, two standards were developed:
  - ISO 19115: has been approved – it is an abstract standard that specifies general content for the metadata, but does not specify the format.
  - ISO 19139: XML implementation schema specifying the metadata record format.



# International Organization for Standardization (ISO)

The ISO allows profiles to be developed for groups sharing a common interest. The FGDC has teamed with Canada to produce the North American Profile (NAP).

The NAP will replace the current Version 2 FGDC Standard, and essentially become Version 3

## ISO 19115 Core Metadata Elements

### Mandatory Elements:

Dataset title  
Dataset reference date  
Dataset language  
Dataset topic category  
Abstract  
Metadata point of contact  
Metadata date stamp

### Conditional Elements:

Dataset responsible party  
Geographic location by coordinates  
Dataset character set  
Spatial resolution  
Distribution format  
Spatial representation type  
Reference system  
Lineage statement  
On-line Resource  
Metadata file identifier  
Metadata standard name  
Metadata standard version  
Metadata language  
Metadata character set

<http://www.fgdc.gov/metadata/geospatial-metadata-standards>

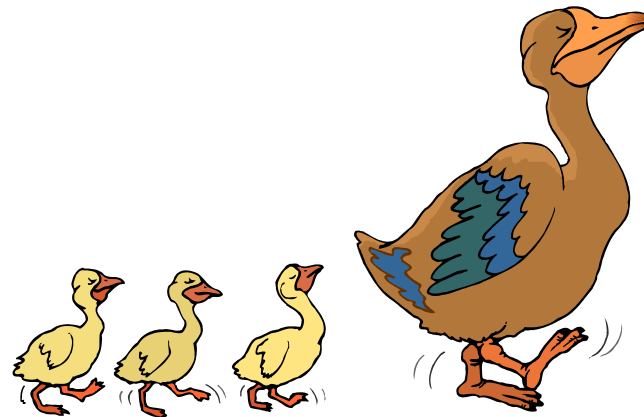
# ISO 19115

- ISO 19115 enables describing:
  - Geospatial data sets
  - Maps
  - Non-geospatial resources (example: tabular data)
  - Services: portals, web mapping



---

# Metadata Clearinghouses



---

# Topics

- Define Clearinghouse
- Examples
- How to use the USGS NBII Clearinghouse

# Metadata Clearinghouses

- A metadata clearinghouse is a location — typically accessed through the Internet — to search for spatial data sets
- Clearinghouses make metadata records easy to find



# Examples of Clearinghouses:

## USGS NBII Clearinghouse

<http://metadata.nbii.gov/clearinghouse>

The screenshot shows the top of the NBII Metadata Clearinghouse website. At the top, there is a logo for 'nbii National Biological Information Infrastructure' and 'NBII Metadata Clearinghouse | HELP'. Below this, there are two tabs: 'Simple Search' and 'Advanced Search'. The main search area is titled 'Search All Records For' and contains a search input field, a 'SEARCH' button, and a 'Results/Page' dropdown menu set to '10'. A hint below the search field reads: 'Hint: boolean operators, wildcards and phrases are allowed. ex: precipitation or (rain\* and "moisture content")'. Below the search area, there is a section for 'Query being built:' which is currently empty and marked as 'Not Editable'. At the bottom left of the search area, there is a 'CLEAR QUERY' link.


[USGS](#) [DOI](#) [oml](#) [HERBERT](#)  
[My NBII](#) | [NBII Disclaimer and Privacy Statement](#) | [Feedback Form](#) | [Email Us](#) | [Security Warning](#)

The screenshot shows the geodata.gov website. The header includes the logo 'geodata.gov U.S. MAPS & DATA' and the tagline 'Your One Stop for Federal, State & Local Geographic Data'. Below the header is a navigation menu with links for Home, Search, Maps, Marketplace, Communities, Statistics, and Help Center. The main content area is divided into several sections: 'Communities' with a search bar and 'Show Advanced Search Options' link; 'Special Interest' with links to various topics like Hurricanes, Fire Mapping, and GIS; 'Data Categories' with links to Administrative Boundaries, Agriculture, Atmosphere, Biology, Business, Cadastral, and Demographic; 'Earth Information Exchange Gateway' with a search bar and 'Show Advanced Search Options' link; and 'Quick Start' with a 'Welcome to geodata.gov' message and a list of links: 'Find Data or Map Services', 'Make a Map', 'Browse Community Information', and 'Cooperate on Data Acquisitions'. At the bottom, there is a section for 'Save searches, maps, and your favorite geography to re-use later' and a 'Quick Start Guide' link.

## GeoSpatial One Stop

<http://gos2.geodata.gov>

# Searching the USGS NBII Clearinghouse:



**nbii** National Biological Information Infrastructure  
NBII Metadata Clearinghouse | [HELP](#) [ABOUT](#)

[Simple Search](#) **Advanced Search** [Map Search](#)

The NBII Clearinghouse is a powerful resource for scientists, allowing them to share and access information about important research in natural resources. With over 90,000 records from over 80 data providers, the possibilities for collaborations and data exchange are endless. Click [here](#) to read more about the NBII Clearinghouse

**Search All Records For**

[SEARCH](#)

*Hint: boolean operators, wildcards and phrases are allowed.  
ex: precipitation or (rain\* and "moisture content")*

Results/Page  
10

**Query being built:**

*Not Editable*

[CLEAR QUERY](#)

# Advanced Search

Simple Search

Advanced Search

## Fielded Search

FullText  OR

FullText  OR

FullText  OR

[Add more options](#)

[Help](#) | [clear](#)

## Date Search

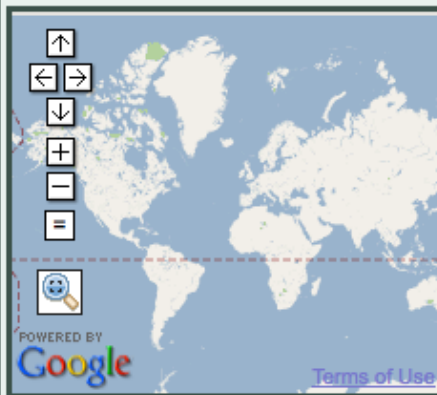
Data Collection Date  
 Publication Date  
 Either

during  thru

mm/dd/yyyy mm/dd/yyyy

[Help](#) | [clear](#)

## Geographic Search



List Areas in:  
USA  WORLD

Select from list

Search Area:  
overlaps  encloses

North

West East

South

\*Click on to select an area

Place:  [view on map](#)

[Help](#) | [clear](#)

## Format

All  
Maps and Data  
Publications  
Tools and Software

## Data providers

All  
**U.S Geological Survey (USGS)**  
BRD Bibliographic Metadata Node  
Columbia Environmental Research Center Metadata Node  
Fire Research and Management Exchange System (FRAMES)  
Forest and Rangeland Ecosystem Science Center Metadata Node  
Great Basin Information Project  
Hawaii Biodiversity and Mapping Program  
National Gap Analysis Program Metadata Node  
National Wetlands Research Center

## Query being built:

Not Editable

[CLEAR QUERY](#)

[SEARCH](#)

# Results



National Biological Information Infrastructure

## Metadata Summary

Email

Bookmark

RSS  
Feed

Help

Your search found: 28 documents.

You searched for: fullText:cattle and ranch AND datasource:( all )

Filter by resource type	Filter by decade	Filter by originator
<a href="#">Maps and Data (19)</a> <a href="#">others (9)</a>	<a href="#">2000 (15)</a> <a href="#">1990 (13)</a>	<a href="#">Gosz, James (10)</a> <a href="#">Moore, Doug (8)</a> <a href="#">usgs (6)</a> <a href="#">Parmenter, Robert (2)</a> <a href="#">Principal investigator(s) (1)</a>

Viewing Documents 1 - 10 out of 28

Prev 1 2 3 Next

Return to Search

Sort By:

Index Rank

Pubdate

Source

### Filter by data providers

[Long Term Ecological Research \(LTER\) Network \(20\)](#)

[Metadata Clearinghouse](#)

ELK AND BISON GRAZING ECOLOGY AND MANAGEMENT IN GREAT SAND DUNES NATIONAL PARK AND PRESERVE

10/01/2004 - 05/31/2008

**Data provider:** METADATA CLEARINGHOUSE PRINCIPAL NODE

The establishment of Great Sand Dunes National Park and Preserve and the new Baca National Wildlife Refuge in the San Luis Valley

# Results

Viewing Documents 1 - 10 out of 28

[Prev](#) [1](#) [2](#) [3](#) [Next](#)

[Return to Search](#)

Sort By:

Index Rank

Pubdate

Source

## Filter by data providers

[Long Term Ecological Research \(LTER\) Network \(20\)](#)

[Metadata Clearinghouse Principal Node \(6\)](#)

[Baruch Coastal Monitoring \(1\)](#)

[California Environmental Information Catalog \(1\)](#)

### ELK AND BISON GRAZING ECOLOGY AND MANAGEMENT IN GREAT SAND DUNES NATIONAL PARK AND PRESERVE

10/01/2004 - 05/31/2008

**Data provider:** METADATA CLEARINGHOUSE PRINCIPAL NODE

The establishment of Great Sand Dunes National Park and Preserve and the new Baca National Wildlife Refuge in the San Luis Valley of Colorado was one of the most significant land conservation actions in the U.S. west in recent years. The action was a result of cooperation between the NPS, FWS, BLM, USFS, and The Nature Conservancy. The new national park will consist of 107,265 acres, the new national preserve, 41,872 acres, and the new national wildlife refuge, 92,180 acres. The area encompassed by this designation protects a number of natural wonders and features including a unique ecosystem o



[View full metadata](#)

### STATE OF CALIFORNIA GRAZING ALLOTMENTS

01/01/2003 - 12/31/2003

**Data provider:** CALIFORNIA ENVIRONMENTAL INFORMATION CATALOG

BLM grazing allotments in California.



[View full metadata](#)

### ELK AND BISON GRAZING ECOLOGY AND MANAGEMENT IN GREAT SAND DUNES NATIONAL PARK AND PRESERVE, ELK BODY CONDITION AND LOCATION DATA FROM 2005 TO 2008.

01/17/2005 - 05/08/2007

**Data provider:** METADATA CLEARINGHOUSE PRINCIPAL NODE

The establishment of Great Sand Dunes National Park and Preserve (GRSA) and the new Baca National Wildlife Refuge in the San Luis Valley of Colorado was one of the most significant land conservation actions in the U.S. west in recent years. The action was a result of cooperation between the NPS, FWS, BLM, USFS, and The Nature Conservancy. The new national park will consist of 107,265 acres, the new national preserve, 41,872 acres, and the new national wildlife refuge, 92,180 acres. The area encompassed by this designation protects a number of natural wonders and features including a unique ecos

# Results



National Biological Information Infrastructure



## Metadata Report

[Bookmark](#) [Email](#)

You searched for: **fulltext:cattle and ranch and datasource:( all )**

[Return to Search](#)

[Return to Results](#)

[View this page in FGDC](#)

**Title:** Elk and bison grazing ecology and management in Great Sand Dunes National Park and Preserve

**Investigator(s):** U.S. Geological Survey, Fort Collins Science Center (FORT)

**Status:** In work

**Access Restriction:** Contact principal investigator after completion of project

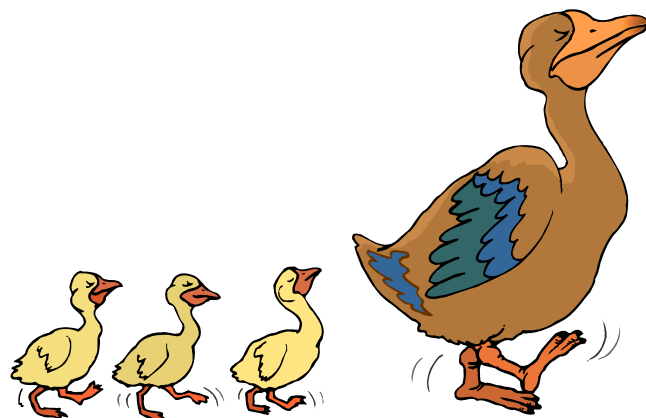
**Data Center Contact:** Phone: Email:

Keyword(s):	Theme Keyword Thesaurus	Theme Keyword
	None	ecology management ELK bison

**Abstract:** The establishment of Great Sand Dunes National Park and Preserve and the new Baca National Wildlife Refuge in the San Luis Valley of Colorado was one of the most significant land conservation actions in the U.S. west in recent years. The action was a result of cooperation between the NPS, FWS, BLM, USFS, and The Nature Conservancy. The new national park will consist of 107,265 acres, the new national preserve, 41,872 acres, and the new national wildlife refuge, 92,180 acres. The area encompassed by this designation protects a number of natural wonders and features including a unique ecosystem of natural sand dunes, the entire

---

# How to write quality metadata



---

## Topics

- Steps for preparing quality metadata
- Tips for writing good metadata

## Steps to Create Quality Metadata

- Organize your information
- Write your metadata
- Review for accuracy and completeness
- Have someone else read your file
- Revise it, based on comments from your reviewer
- Review it once more before you publish it



## Tips for Writing Good Metadata

Think about the long-term effects:

- Do not use jargon
- Define technical terms and acronyms:  
CA, LA, GPS, GIS
- Clearly state data limitations
- Use “none” or “unknown” meaningfully
  - None usually means that you knew about data and nothing existed; (e.g., a “0” cubic feet per second discharge value)
  - Unknown means that you don’t know whether that data existed or not. (e.g., null value)



---

## Tips for Writing Good Metadata

### *Titles, Titles, Titles...*

- Titles are *critical* in helping readers find your data
  - While individuals are searching for the most appropriate datasets on the clearinghouses, they are most likely going to use the title as the first criteria to determine if a dataset meets their needs.
  - Treat the title as the opportunity to sell your dataset.
- A *complete title* includes: *What, Where, When, Who, and Scale*
- An *informative title* includes: *topic, timeliness of the data, specific information about place and geography*

## Tips for Writing Good Metadata

- Which title is better?

- *Rivers*

*OR*

- *Greater Yellowstone Rivers from 1:126,700  
U.S. Forest Service Visitor Maps (1961-1983)*



---

*Greater Yellowstone (where) Rivers (what) from  
1:126,700 (scale) U.S. Forest Service (who)  
Visitor Maps (1961-1983) (when)*

---

---

## Tips for Writing Good Metadata

*Be specific and quantify when you can:* the goal of a metadata record is to give the user enough information to know if they can use it without contacting the dataset owner.

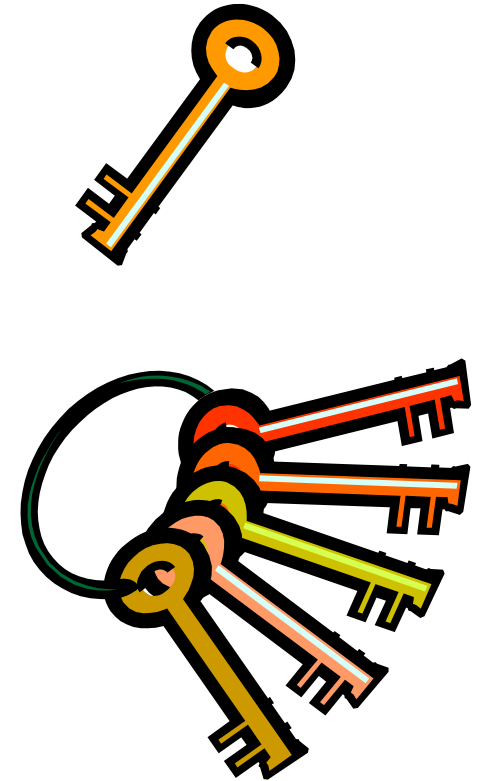
**Vague:** *We checked our work and it looks complete.*

**Specific:** *We checked our work using 3 separate sets of check plots reviewed by 2 different people. We determined our work to be 95% complete based on these visual inspections.*

# Tips for Writing Good Metadata

## *Select keywords wisely*

- Use unambiguous words
- Use descriptive words
- Fully qualify geographic locations
- Use thesauri whenever possible
  - Example: USGS NBII Biocomplexity Thesaurus (over 9,500 terms (<http://thesaurus.nbii.gov>))



# Tips for Writing Good Metadata

*Remember: a computer will read your metadata*

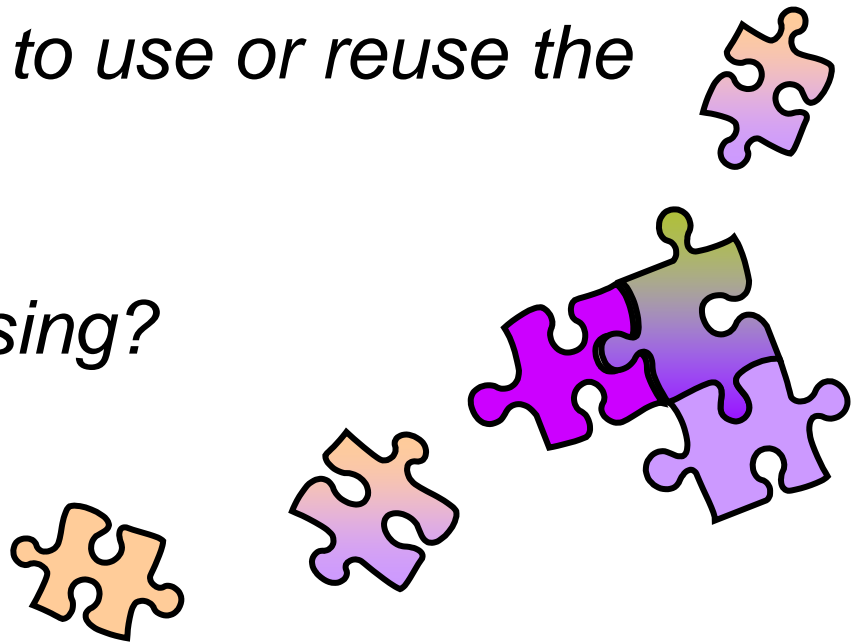
- Don't use symbols that might be misinterpreted
  - ! @ # % { } | / \ < > ~
- Don't use characters with dual interpretations
  - ie: < > signs are used in html coding
- Don't use tabs or indents
- When copying and pasting from other sources, use a text editor (e.g., Notepad) to eliminate hidden characters



## Tips for Writing Good Metadata

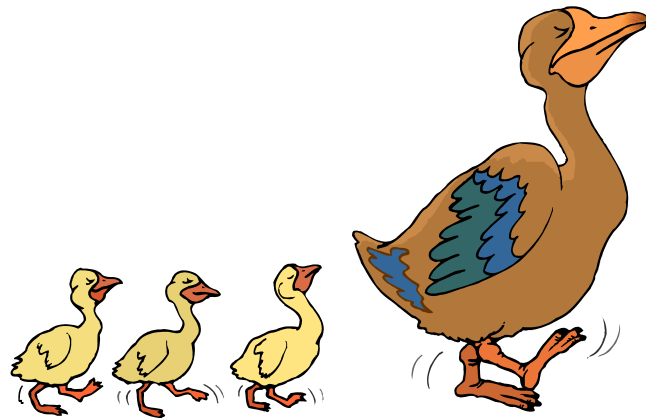
*\*\*Review your final product\*\**

- *Does the documentation present all the information needed to use or reuse the data?*
- *Are any pieces missing?*



---

# Metadata Tools and Resources



# Metadata Creation Tool Reviews

<b>Software Name</b>	<b>Spatial Capture</b>	<b>Templates</b>	<b>BDP</b>	<b>Price</b>
ArcCatalog (ArcGIS 9.x)	YES	Import existing metadata, Advanced Synchronization tools	NO	Part of ArcGIS 9.x
Metavist	NO	Stores as XML files; can start with existing metadata to use as a template	YES	Free
SMMS	YES	YES – creates whole record templates, and citations & contacts	YES	~ \$600
TKME	NO	Start with existing metadata record to use as a template	YES	Free

# ESRI's ArcCatalog

The image displays two overlapping dialog boxes from the ArcCatalog software. The background dialog is titled "Editing 'NatureServe Element Occurrence Data for Rare and Endangered Species in the United States and Canada...". It features a tabbed interface with "Identification" selected. The "Description" tab is active, showing fields for "Abstract", "Purpose", "Language", and "Supplemental Information". The "Abstract" field contains the text: "NatureServe, in collaboration with its member Natural Heritage Programs and Conservation Data Centres, maintains a database of rare and imperiled species and plant communities across the United States and Canada. The Element". The "Purpose" field contains: "The Element Occurrence (EO) data are collected and maintained by the Natural Heritage Programs and Conservation Data Centres using a standardized methodology to provide accurate and current biodiversity and conservation". Below these are fields for "Access Constraints" and "Use Constraints". The "Access Constraints" field contains: "Heritage information is only for the intended use of the individual or organization who requested it. This database and accompanying files may not be distributed in any way without the consent of NatureServe. Use of the data is subject to". The "Use Constraints" field contains: "Use of the Element Occurrence Dataset by the licensed data users shall be governed by the following provisions:". There are also fields for "Data Set Credit" and "Native Data Set Environment". The "Data Set Credit" field contains: "NatureServe aggregates data developed and maintained by the network of Natural Heritage Programs and Conservation Data Centres, which is made up of 75 independent centers in the U.S., Canada, and Latin America." The "Native Data Set Environment" field contains: "NatureServe's databases are managed in Biotics, a custom Oracle-based application that incorporates a custom GIS interface in ArcView 3.2. Biotics is a customized information management system designed specifically for the Natural". At the bottom of this dialog are "Save", "Cancel", and "Help" buttons.

The foreground dialog is also titled "Editing 'NatureServe Element Occurrence Data for Rare and Endangered Species in the United States and Canada...". It has tabs for "Detailed Description" and "Overview Description", with "Detailed Description" selected. It has sub-tabs for "Entity Type" and "Attribute", with "Attribute" selected. Under "Attribute", there are sub-tabs for "General", "Dates", and "Attribute Domain Values", with "General" selected. The "Label" field contains "GRANK". The "Type" field is empty. The "Width" field is empty. The "Precision" field is empty. The "Indexed" field is empty. The "Definition" field contains "Global Rank - The NatureServe Conserv". The "Definition Source" field contains "NatureServe - Heritage methodology". At the bottom of this dialog are "Save", "Cancel", and "Help" buttons. The status bar at the bottom of the foreground dialog shows "Attribute 28 of 65" and "Detailed Description 1 of 1".

# ArcCatalog with NPS NBII Extension

**ArcCatalog - ArcInfo - C:\gis\DATA\bndlostrail**

File Edit View Go Tools Help

Stylesheet: FGDC NBII

Catalog  
C:\  
96fc8f1ebb3bd0ebdefdbd8678d  
arcgis  
burnseverity  
d7b59  
data backup  
dell  
DianeTmp  
Documents and Settings  
ESRI  
Forms  
gis  
ArtAllen  
AssignmentSheets  
burn\_severity  
CWD  
DATA  
bndlostrail  
cogotest  
county\_1a  
county\_dd  
r6stat\_e  
r6stations  
states100k  
statesp020  
armi.info  
citiesx020.shp  
CO\_cities.shp  
condivl020.shp  
county\_1a.shp  
county\_1ap.txt  
lrbnd.e00  
westernus.shp  
dem  
Digitizing  
drgs  
Erin

Contents Preview Metadata

**Maintenance and update frequency:** Continually

**Spatial domain:**  
**Description of geographic extent:**  
Portions of the Upper Missouri and Yellowstone r... study area was divided into 3 segments - Segme... of the Yellowstone River, from the Missouri conf... Intake. Segment 2 was defined as the Wild and... the headwaters of Fort Peck Reservoir (rkm432)... encompassed the reach downstream from Fort P...

**Bounding coordinates:**  
\* **West bounding coordinate:** -114.977167  
\* **East bounding coordinate:** -114.785509  
\* **North bounding coordinate:** 48.207435  
\* **South bounding coordinate:** 48.146351

**Local bounding coordinates:**  
\* **Left bounding coordinate:** 650458.119912  
\* **Right bounding coordinate:** 664539.598407  
\* **Top bounding coordinate:** 5341120.111770  
\* **Bottom bounding coordinate:** 5334714.113438

**Keywords:**  
**Theme:**  
**Theme keywords:** Decision Support System, Two dimensional modeling, warm-water fishes, reservoir operations, hydrology  
**Theme keyword thesaurus:** None

**Place:**  
**Place keywords:** Montana, MT, North Dakota, ND, Yellowstone River, Missouri River  
**Place keyword thesaurus:** None

**Taxonomy:**  
**Keywords/taxon:**  
**Taxonomic keywords:** warmwater fishes  
**Taxonomic keyword thesaurus:** none

NPS Metadata  
Import E00  
Import Metadata  
Import XML Template  
Export E00  
Export Metadata  
Parse with MP  
Metadata Template  
Spellcheck Metadata  
Metadata Catalog  
Metadata Search  
Help

# Metavist

The image shows a screenshot of the Metavist software interface. The main window is titled "Metavist - natsrv\_d\_eo.xml" and has a menu bar with "File", "Options", and "Help". Below the menu bar is a toolbar with icons for file operations and a set of tabs: "Identification", "Data Quality", "Spatial Data Org", "Spatial Reference", "Entity & Attribute", "Distribution", and "Metadata Ref". Underneath the tabs is another set of icons for "Basic Info", "Spatial Domain", "Keywords", "Taxonomy", "Access", "Analytical Tools", and "Miscellaneous".

The main content area is divided into sections: "Citation for the data" with fields for "Author(s)" (containing "Lynn Kutner (ed.), NatureServe") and "Title" (containing "NatureServe Element Occurrence Data for Rare and Endangered Species"); "Abstract" (containing a paragraph about NatureServe's database); and "Purpose" (containing a paragraph about data collection). A "Contact Information" dialog box is overlaid on the right side of the main window.

The "Contact Information" dialog box has a title bar and a close button. It contains the following fields and controls:

- The primary contact is a(n):** Radio buttons for "Person" (selected) and "Organization".
- Contact Person:** Text field containing "Shara Howie".
- Contact Organization (optional):** Text field containing "NatureServe".
- Contact Position (optional):** Text field containing "Director, Heritage Data Services".
- Contact Address:** Text field containing "mailing and physical address: Arlington". Below this are three buttons: "Add", "Edit", and "Delete".
- Contact Voice Telephone:** Text field containing "(703) 908-1855".
- Contact TDD/TTY Telephone (optional):** Text field.
- Contact Fax Telephone (optional):** Text field containing "(703) 908-1917".
- Contact E-Mail Address (optional):** Text field containing "shara\_howie@natureserve.org".
- Hours of Service (optional):** Text field.
- Contact Instructions (optional):** Text field.

At the bottom right of the dialog box are "OK" and "Cancel" buttons.

# Spatial Metadata Management System (SMMS)

SMMS 3.2 - [C:\LynnStuff\Work\metadata\NS\_metadata.mdb]

File Edit View Tools Help

Title: NatureServe Element Occurrence Data for Rare and Endangered Species in the United States and Canada

General | Time Period | Status | Bounding Coord. | G-Polygon | Keywords | Tax. System | Tax. Class. | Browse Graphic | Security | Cross Ref. | Tools

Citation: NatureServe Element Occurrence Data for Rare and Endangered Species in the United States and Canada

Point of Contact: Shara Howie

Description

Abstract: NatureServe, in collaboration with its member Natural Heritage Programs and Conservation Data Centres, maintains a database of rare and imperiled species and plant communities across the United States and Canada. The Element

Purpose: The Element Occurrence (EO) data are collected and maintained by the Natural Heritage Programs and Conservation Data Centres using a standardized methodology to provide accurate and current biodiversity and conservation information to

Supplemental Information

Access Constraints: Heritage information is only for the intended use of the individual or organization who requested it. This database and accompanying files may not be distributed in any way without the consent of NatureServe. Use of the data is subject to the

Use Constraints: Use of the Element Occurrence Dataset by the licensed data users shall be governed by the following provisions:

Data Set Credit: NatureServe aggregates data developed and maintained by the network of Natural Heritage Programs and Conservation Data Centres, which is made up of 75 independent centers in the U.S., Canada, and Latin America.

Native Data Set Environment: NatureServe's databases are managed in Biotics, a custom Oracle-based application that incorporates a custom GIS interface in ArcView 3.2. Biotics is a customized information management system designed specifically for the Natural Heritage Network.

Identification | Data Quality | Spatial Data Organization | Spatial Reference | Entity and Attribute | Distribution | Metadata Reference | Associated Data | User Defined

Editing Metadata, Press Shift-F1 for help on current field.

Entity Definition: Locational data for species and ecological communities, as defined by NatureServe Heritage methodology.

Entity Definition Source: NatureServe - Heritage methodology

NUM

Detailed Description 1 of 1

Attribute Labels

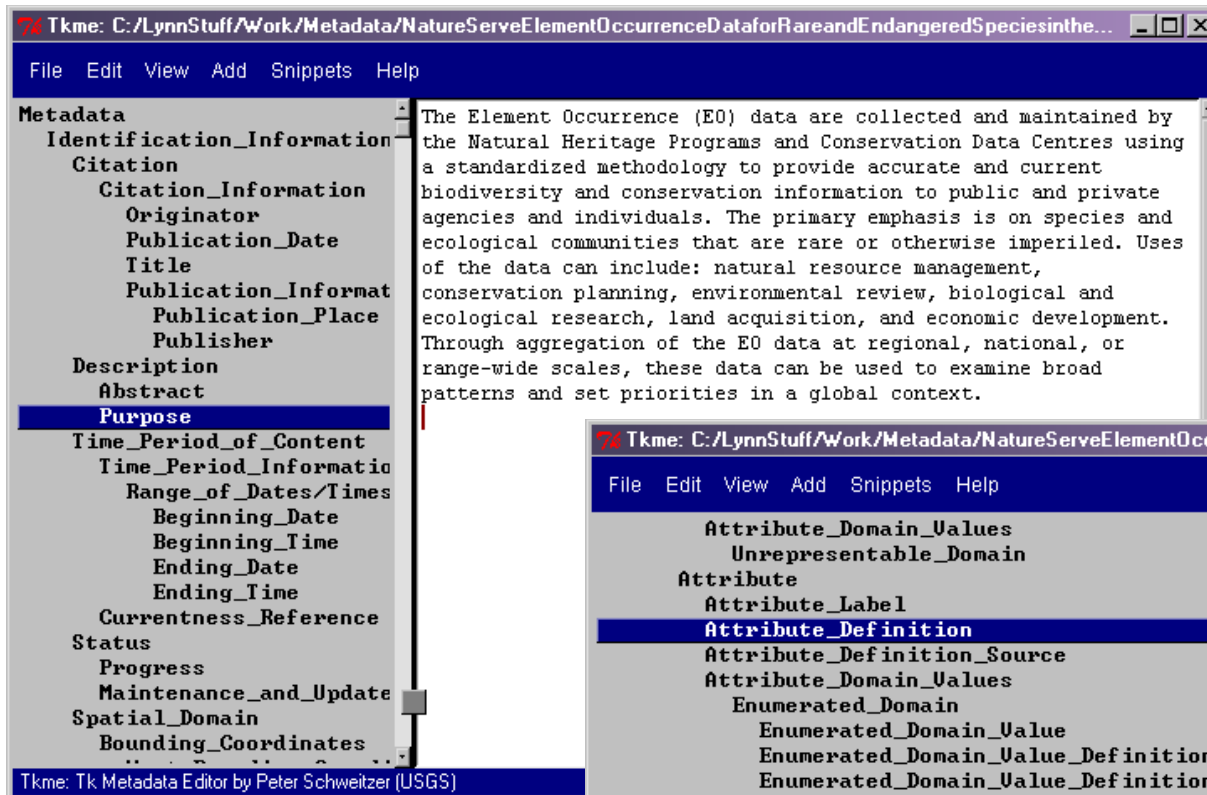
- Biotics Mapper: CONF\_EXT
- Biotics Mapper: EST\_REP\_AC
- Biotics Mapper: MULTI\_JUR
- Biotics Mapper: REP\_AC\_CDM
- Biotics Mapper: SEP\_CDM
- Biotics Mapper: SEP\_DIST
- CITATIONS
- CLASS\_STAT

7 of 65

Identification | Data Quality | Spatial Data Organization | Spatial Reference | Entity and Attribute | Distribution | Metadata Reference | Associated Data | User Defined

Editing Metadata, Press Shift-F1 for help on current field.

# USGS TKME “Another editor for formal metadata”



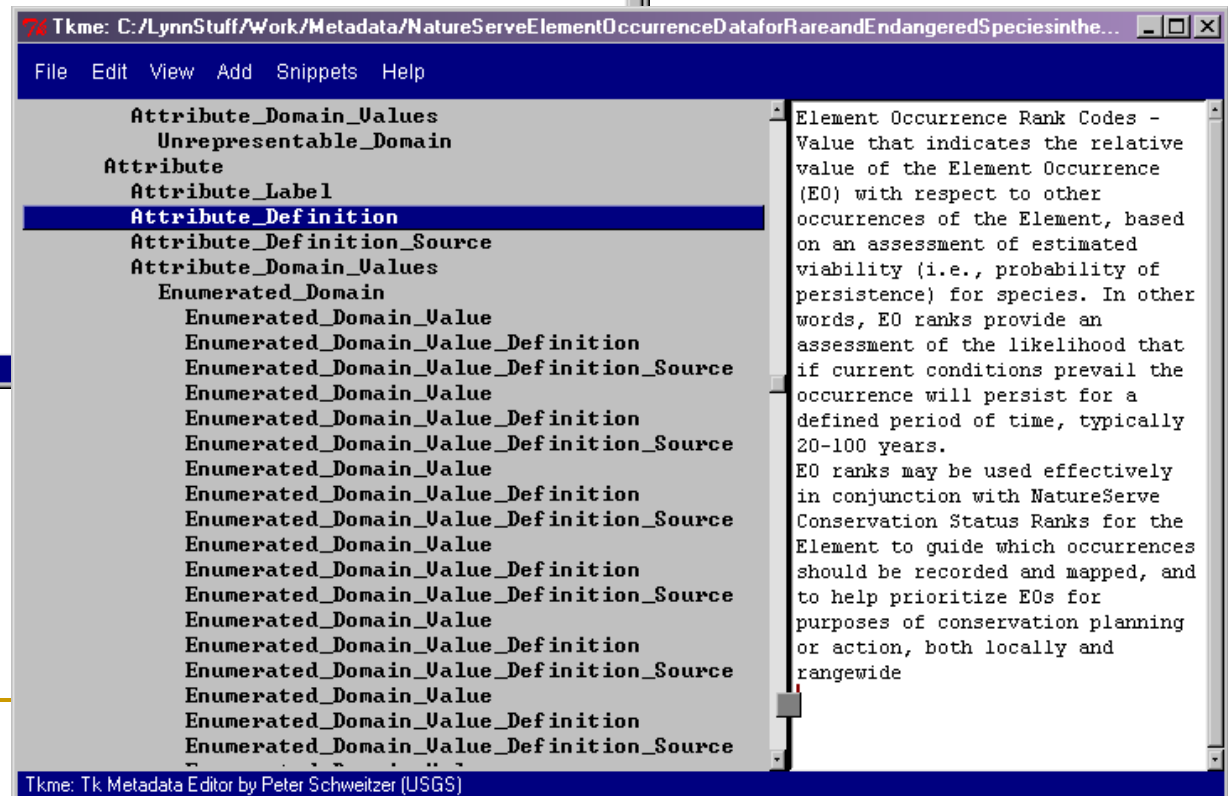
The screenshot shows the Tkme Metadata Editor interface. On the left is a tree view of metadata elements, with 'Purpose' selected. The main text area on the right contains the following text:

The Element Occurrence (EO) data are collected and maintained by the Natural Heritage Programs and Conservation Data Centres using a standardized methodology to provide accurate and current biodiversity and conservation information to public and private agencies and individuals. The primary emphasis is on species and ecological communities that are rare or otherwise imperiled. Uses of the data can include: natural resource management, conservation planning, environmental review, biological and ecological research, land acquisition, and economic development. Through aggregation of the EO data at regional, national, or range-wide scales, these data can be used to examine broad patterns and set priorities in a global context.

The tree view on the left includes the following categories:

- Metadata
  - Identification\_Information
    - Citation
      - Citation\_Information
        - Originator
        - Publication\_Date
        - Title
        - Publication\_Information
        - Publication\_Place
        - Publisher
  - Description
    - Abstract
  - Purpose**
  - Time\_Period\_of\_Content
    - Time\_Period\_Information
      - Range\_of\_Dates/Times
        - Beginning\_Date
        - Beginning\_Time
        - Ending\_Date
        - Ending\_Time
  - Currentness\_Reference
  - Status
    - Progress
    - Maintenance\_and\_Update
  - Spatial\_Domain
    - Bounding\_Coordinates

Tkme: Tk Metadata Editor by Peter Schweitzer (USGS)



The screenshot shows the Tkme Metadata Editor interface with 'Attribute\_Definition' selected in the tree view. The main text area on the right contains the following text:

Element Occurrence Rank Codes - Value that indicates the relative value of the Element Occurrence (EO) with respect to other occurrences of the Element, based on an assessment of estimated viability (i.e., probability of persistence) for species. In other words, EO ranks provide an assessment of the likelihood that if current conditions prevail the occurrence will persist for a defined period of time, typically 20-100 years. EO ranks may be used effectively in conjunction with NatureServe Conservation Status Ranks for the Element to guide which occurrences should be recorded and mapped, and to help prioritize EOs for purposes of conservation planning or action, both locally and rangewide

The tree view on the left includes the following categories:

- Attribute\_Domain\_Values
  - Unrepresentable\_Domain
- Attribute
  - Attribute\_Label
  - Attribute\_Definition**
  - Attribute\_Definition\_Source
  - Attribute\_Domain\_Values
    - Enumerated\_Domain
    - Enumerated\_Domain\_Value
    - Enumerated\_Domain\_Value\_Definition
    - Enumerated\_Domain\_Value\_Definition\_Source
    - Enumerated\_Domain\_Value
    - Enumerated\_Domain\_Value\_Definition
    - Enumerated\_Domain\_Value\_Definition\_Source
    - Enumerated\_Domain\_Value
    - Enumerated\_Domain\_Value\_Definition
    - Enumerated\_Domain\_Value\_Definition\_Source
    - Enumerated\_Domain\_Value
    - Enumerated\_Domain\_Value\_Definition
    - Enumerated\_Domain\_Value\_Definition\_Source
    - Enumerated\_Domain\_Value
    - Enumerated\_Domain\_Value\_Definition
    - Enumerated\_Domain\_Value\_Definition\_Source
    - Enumerated\_Domain\_Value
    - Enumerated\_Domain\_Value\_Definition
    - Enumerated\_Domain\_Value\_Definition\_Source

Tkme: Tk Metadata Editor by Peter Schweitzer (USGS)

# NOAA Mermaid: An online tool



NOAA Satellite and Information Service  
National Environmental Satellite, Data, and Information Service (NESDIS)



National Coastal Data  
Development Center

home » mermaid metadata resources » tools

Home

About Us

Metadata Resources

Regional Offices

Projects

training

**tools**

additional information

references

## Metadata Enterprise Resource Management Aid (MERMAid)



### Getting Started with MERMAid

- [Request an Account](#)
- [V1.2 Getting Started Guide](#)  
(PDF 5 MB)
- [V1.2 Getting Started Guide](#)

NCDDC provides coastal data resources (organizations and individuals) with a tool to develop, validate, manage and publish metadata records via secure internet access. The Metadata Enterprise Resource Management Aid (MERMAid) allows users/data providers to establish unlimited metadata databases to organize their metadata records any way they see fit (i.e. by program, project, data type, personnel). Some of the key features in MERMAid include (1) user-defined roles and permissions at the metadata management and database levels; (2) change tracking; and (3) enhanced validation. Also, your existing FGDC compliant metadata (in XML format) can be ingested into and managed through MERMAid.

In the near future, NCDDC will be shifting from its current metadata catalog to a knowledge base catalog. MERMAid will play an integral role in this transition. To better leverage these new capabilities, enhanced search and discovery tools will be made available to the public and metadata managers that will provide powerful drill-down features.

### NCDDC Services

[Site Map](#)  
[Metadata Search](#)  
[Middleware Technology](#)  
[Download](#)  
[Interactive Maps](#)  
[Regional Ecosystems](#)  
[C-SIDE](#)

### NOAA Services

[Central Library](#)  
[Photo Library](#)  
[Video Library](#)  
[Visualization Lab](#)  
[Education Resources](#)  
[National Locator](#)  
[NOAA In Your State](#)

### SSC Visitor Services

[Regional Map](#)  
[Highway Map](#)  
[Additional Information](#)

# EPA Metadata Editor

EPA Metadata Editor

Basic Data Set Information | Quality, Coordinate System, and Attribute Information | Distribution & Metadata Information

**Citation**

\* **Origin:** United States Environmental Protection Agency, Office of Environment default

\* **Title:** EPAs Web Feature Service for the Facility Registry System

**Publisher**

\* **Published By:** U.S. Environmental Protection Agency default

\* **Published at:** Washington, DC \* **Date:** 2006 today

**Online Linkage**

\*\* **Primary Linkage:** <http://intramap33.rtpnc.epa.gov/wfsconnector/com.e> ✓

\*\* **Secondary Linkage:** <http://www.epa.gov/frs/index.htm> ✓

**Description**

\* **Abstract:** This web feature service provides location and attribute information on facilities, sites or places subject to environmental regulations or environmental interest. There are over 1.9 million

\* **Purpose:** The purpose of this web feature service is to provide users with access to information stored in EPA's Facility Registry System (FRS). The information in FRS may be used to track sites,

**Supplemental Info:** This is an EPA-Produced Data Set. This data set is one in a s

**Time Period** default

\* **Date:** 2006 today \* **What is data current to:** publication date

\* **Progress:** In work \* **How often is data updated?:** Monthly

**Bounding Box**

N:71.490303 E:-67.434525 S:17.30E default

\* **N:** 71.490303 \* **E:** -67.434525

\* **S:** 17.305182 \* **W:** -173.129822

**Keywords**

* ISO default	** EPA default	** Place default
biota	Surface Water	Alabama
boundaries	Toxics	Alaska
climatologyMe	Transportatic	American San
economy	Waste	Arizona
elevation	Water	Arkansas
environment	emergency re	California
farming	environmenta	Canada

**Data Set Constraints**

\* **Access:** This service is provided for EPA inter

\* **Use:** None. Please check sources, scale, a

\* **Security Classification:** No Confidentiality (from meta)

**Contact**

Primary Person  Primary Organization

U.S. Environmental Protection Agency | EP

open dbase default

Refresh DB Set Default Spell Check Validate EPA

YELLOW \* mandatory GREEN \*\* mandatory if applicable BLUE optional Click on text to link to element description Save & Close Discard & Close Help

---

# Metadata Creation Tools: Reviews

Summary and review of a variety of metadata software tools:



<http://www.fgdc.gov/metadata/geospatial-metadata-tools>

*Check back periodically – this site is updated on an ongoing basis!*

---

# Metadata Validation Tools

## MP - “Metadata Parser”

- Validation that metadata are FGDC compliant
- Critical if posting metadata to an online clearinghouse
- <http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>

---

# Metadata Standard: Resources

FGDC Metadata Standards (general resource):

<http://www.fgdc.gov/standards/>

Content Standard for Digital Geospatial Metadata (CSDGM)  
(version 2.0), FGDC-STD-001-1998

<http://www.fgdc.gov/metadata/geospatial-metadata-standards>

Content Standard for Digital Geospatial Metadata, Part 1:  
Biological Data Profile, FGDC-STD-001.1-1999

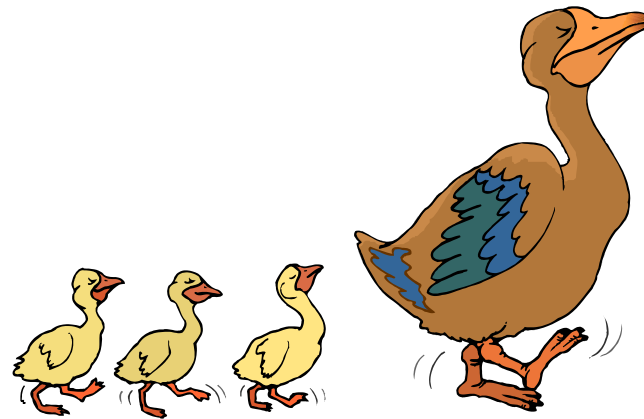
<http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf>

Data Management Toolkit:

<http://pubs.usgs.gov/of/2009/1170/>

---

# The Metadata Process: Step by Step



# To create FGDC compliant metadata records:

**Orient  
yourself with  
this  
workbook**



National Spatial Data Infrastructure

## **Biological Data Profile Workbook**

(For use with FGDC-STD-001-1998)

**Federal Geographic Data Committee**

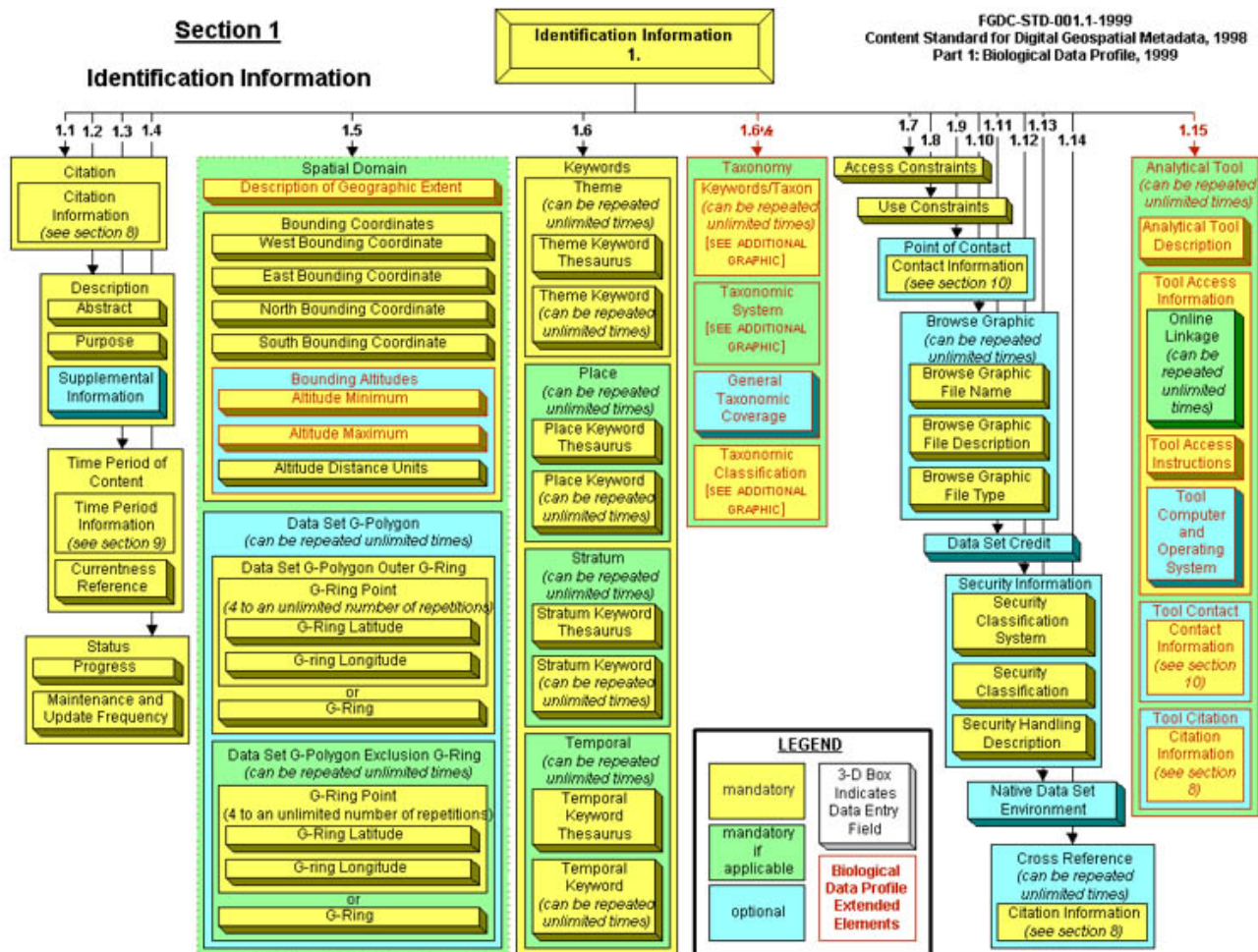
**March 30, 2001**

---

Federal Geographic Data Committee

Department of Agriculture • Department of Commerce • Department of Defense • Department of Energy  
Department of Health & Human Services • Department of Housing and Urban Development  
Department of the Interior • Department of Justice • Department of State  
Department of Transportation • Environmental Protection Agency  
Federal Emergency Management Agency • Library of Congress  
National Aeronautics and Space Administration • National Archives and Records Administration  
National Science Foundation • Tennessee Valley Authority

# Use this graphic to understand the standard



# Address the 7 Sections that make up the Standard:

1. Identification Information
2. Data Quality Information
3. Spatial Data Information
4. Spatial Reference Information
5. Entity and Attribute Information
6. Data Distribution Information
7. Metadata Reference Information



# Use established thesauri

## Browse the Biocomplexity Thesaurus

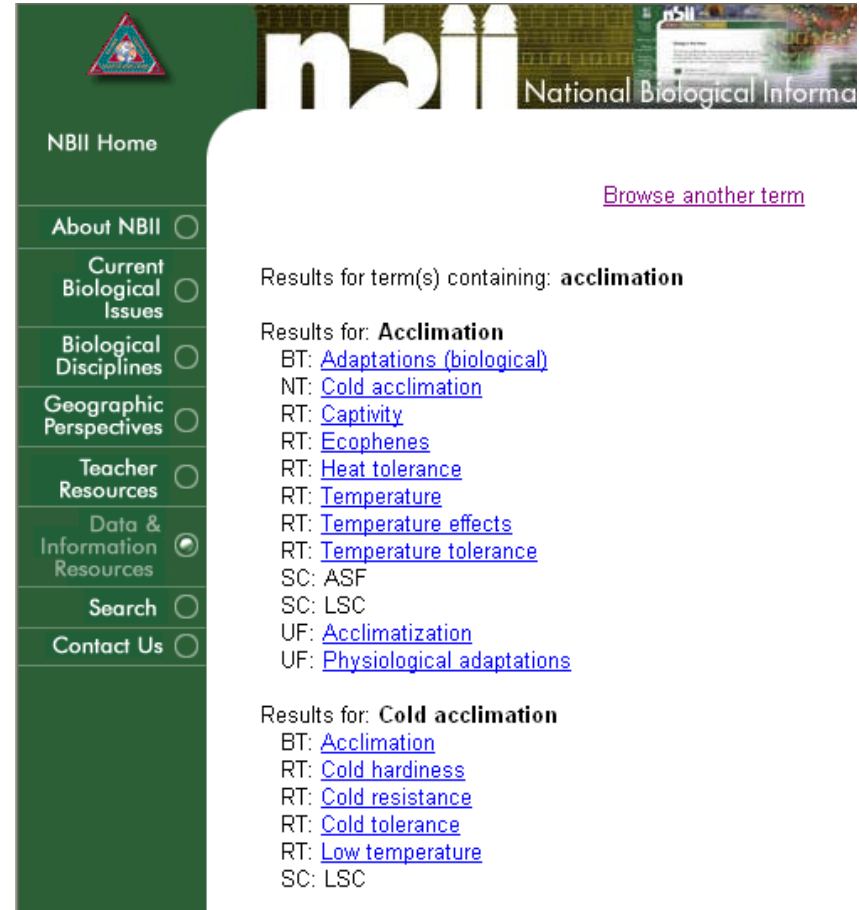
The Biocomplexity Thesaurus was developed through a partnership between the NBII and [CSA](#), a leading bibliographic database provider.

Please enter a term to browse:

Check Thesaurus

- Search on single term or a phrase. (e.g. biodiversity, invasive species)
- Search performs automatic stemming for prefixes and suffixes (e.g. entering [face](#) will yield cell surface, air-ice interface, face, and other variations; entering [bio](#) will yield abiotic factors, biometrics, biodiversity, ectosymbionts, etc.)
- [More about the Biocomplexity Thesaurus](#)

<http://thesaurus.nbii.gov>



The screenshot shows the NBII Home page with a search bar and a list of results for the term 'acclimation'. The results are categorized by type (BT, NT, RT, SC, UF) and include links to related terms.

**Results for term(s) containing: **acclimation****

Results for: **Acclimation**

- BT: [Adaptations \(biological\)](#)
- NT: [Cold acclimation](#)
- RT: [Captiveity](#)
- RT: [Ecophenes](#)
- RT: [Heat tolerance](#)
- RT: [Temperature](#)
- RT: [Temperature effects](#)
- RT: [Temperature tolerance](#)
- SC: ASF
- SC: LSC
- UF: [Acclimatization](#)
- UF: [Physiological adaptations](#)

Results for: **Cold acclimation**

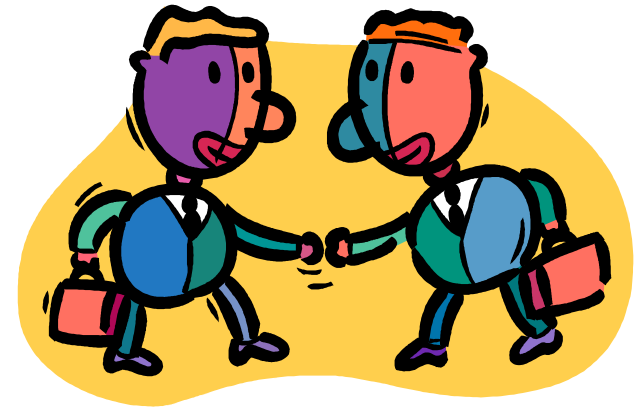
- BT: [Acclimation](#)
- RT: [Cold hardiness](#)
- RT: [Cold resistance](#)
- RT: [Cold tolerance](#)
- RT: [Low temperature](#)
- SC: LSC

# Enter Metadata using an Electronic Tool

The screenshot shows the Metavist software interface. The window title is "Metavist" and it has a menu bar with "File", "Options", and "Help". Below the menu bar is a toolbar with icons for file operations. The main interface is divided into several tabs: "Identification", "Data Quality", "Spatial Data Org", "Spatial Reference", "Entity & Attribute", "Distribution", and "Metadata Ref". Under the "Identification" tab, there are sub-tabs: "Basic Info", "Spatial Domain", "Keywords", "Taxonomy", "Access", "Analytical Tools", and "Miscellaneous". The "Basic Info" sub-tab is active, showing a form for "Citation for the data set". This form includes fields for "Author(s)", "Publication Date", and "Title", along with an "Edit Citation" button. Below the citation form is a large text area for the "Abstract", and another text area for the "Purpose". The interface is designed for data entry and is presented in a yellow-themed window.

# Become a Clearinghouse partner:

- Upload records to a web-accessible folder
- Records should be in .xml format
  - Inform USGS NBII of folder location (*one-time only*)
    - USGS NBII will harvest records weekly



# Search for records in Clearinghouses

Viewing Documents 1 - 10 out of 28

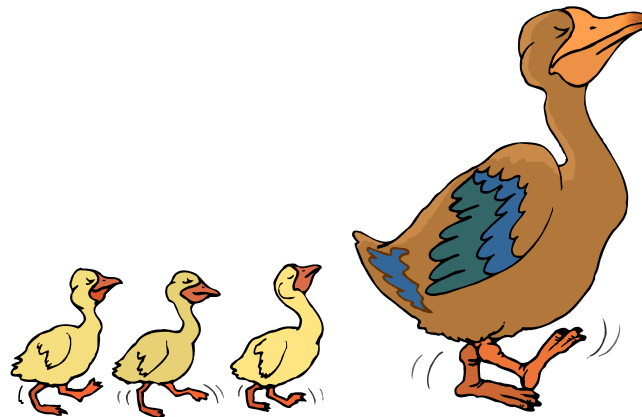
[Prev](#) [1](#) [2](#) [3](#) [Next](#)  
[Return to Search](#)

Sort By: Index Rank Pubdate Source

Filter by data providers	ELK AND BISON GRAZING ECOLOGY AND MANAGEMENT IN GREAT SAND DUNES NATIONAL PARK AND PRESERVE
<a href="#">Long Term Ecological Research (LTER) Network (20)</a> <a href="#">Metadata Clearinghouse Principal Node (6)</a> <a href="#">Bartch Coastal Monitoring (1)</a> <a href="#">California Environmental Information Catalog (1)</a>	10/01/2004 - 05/31/2008
	<p><i>Data provider:</i> METADATA CLEARINGHOUSE PRINCIPAL NODE</p> <p>The establishment of Great Sand Dunes National Park and Preserve and the new Baca National Wildlife Refuge in the San Luis Valley of Colorado was one of the most significant land conservation actions in the U.S. west in recent years. The action was a result of cooperation between the NPS, FWS, BLM, USFS, and The Nature Conservancy. The new national park will consist of 107,265 acres, the new national preserve, 41,872 acres, and the new national wildlife refuge, 92,180 acres. The area encompassed by this designation protects a number of natural wonders and features including a unique ecosystem.</p> <p style="text-align: right;">★★★★★★★★ <a href="#">View full metadata</a></p>
	01/01/2003 - 12/31/2003
	<p><b>STATE OF CALIFORNIA GRAZING ALLOTMENTS</b></p> <p><i>Data provider:</i> CALIFORNIA ENVIRONMENTAL INFORMATION CATALOG</p> <p>BLM grazing allotments in California.</p> <p style="text-align: right;">★★★★★★★★ <a href="#">View full metadata</a></p>
	01/17/2005 - 05/08/2007
	<p><b>ELK AND BISON GRAZING ECOLOGY AND MANAGEMENT IN GREAT SAND DUNES NATIONAL PARK AND PRESERVE, ELK BODY CONDITION AND LOCATION DATA FROM 2005 TO 2008.</b></p> <p><i>Data provider:</i> METADATA CLEARINGHOUSE PRINCIPAL NODE</p> <p>The establishment of Great Sand Dunes National Park and Preserve (GRSA) and the new Baca National Wildlife Refuge in the San Luis Valley of Colorado was one of the most significant land conservation actions in the U.S. west in recent years. The action was a result of cooperation between the NPS, FWS, BLM, USFS, and The Nature Conservancy. The new national park will consist of 107,265 acres, the new national preserve, 41,872 acres, and the new national wildlife refuge, 92,180 acres. The area encompassed by this designation protects a number of natural wonders and features including a unique ecos</p>

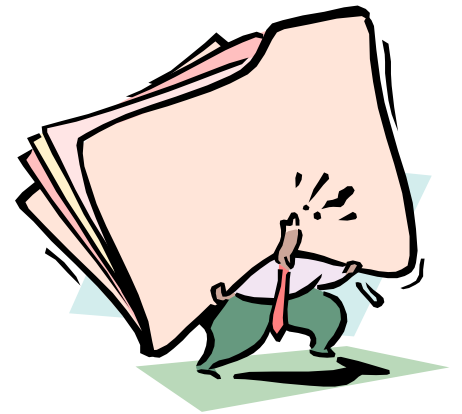
---

# Metadata Implementation



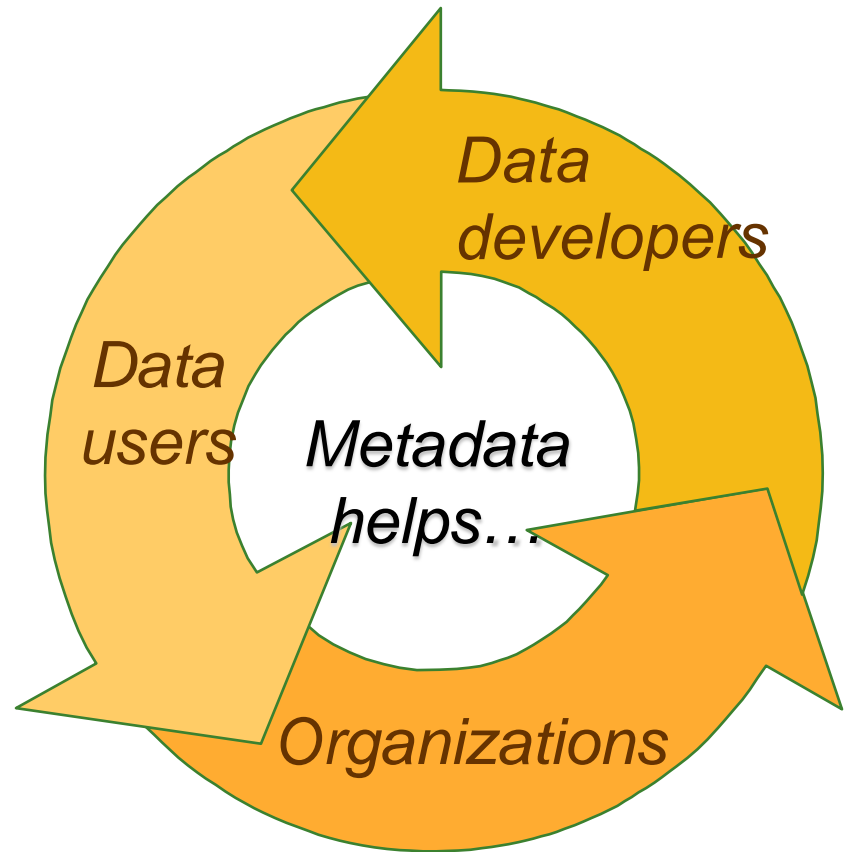
# Topics

- Strategies for metadata implementation
- Tips for prioritizing metadata creation
- Approaches to creating metadata



# Getting Support: The Value of Metadata

- Metadata saves time, money, & frustration
- Preserves institutional memory and investment in data; written documentation rather than in one person's brain
- Partnerships and “advertising” data collections, help share reliable information
- Efficiency – identify and use existing datasets, avoid duplication of effort
- Gives the dataset creator(s) credit



# Funding Metadata Development

## ■ Internal:

- Are there any internal funding options for metadata implementation?

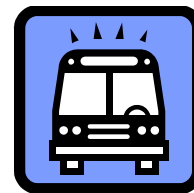
## ■ External:

- Include metadata records as a project deliverable or scope of work when applying for a grant, therefore including it in the project budget.
- CAP Grant Program: The FGDC offers small grants for metadata implementation in several categories. Link: [www.fgdc.gov](http://www.fgdc.gov)



# Deciding what datasets need metadata...

- Prioritize datasets: Mission critical, High demand, Projects, Web downloadable, Legacy, Others?
- Data inventory: Create a quick look up table for your organization that includes Format, Resolution, Contacts, Geographic locations, etc
- General rule of thumb: The “what if Joe Scientist is unable to do his/her job tomorrow?” Will someone be able to pick up where he/she left off?



# Who contributes to creating a metadata file?

*Single individual or team approach?*

- Team Leader / Project Manager
- GIS Specialist
- Field Personnel
- Database Manager
- Science Staff
- Data Analysis Lead



---

# “Working Smart” is the Key

- Create and use metadata templates
- Capitalize on capabilities of various software packages
- Use existing documentation so as to not recreate information
  - Project proposals and notes
- Team approach: experts of specific areas contribute that information
- Example workflows: Data Management Toolkit – Section 9: <http://pubs.usgs.gov/of/2009/1170/>

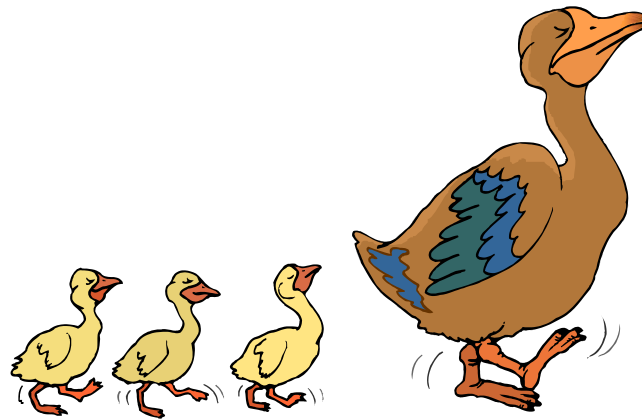
# More Metadata Training & Information Online

- FGDC metadata trainer directory
- FGDC metadata training calendar
  - <http://www.fgdc.gov>
- USGS NBII metadata training program
  - <http://www.nbii.gov>
- Other Training Materials
  - <http://www.csc.noaa.gov/metadata/>
- Email listserv
  - <http://lists.geocomm.com/mailman/listinfo/metadata>



---

# The USGS Biological Informatics Metadata Program

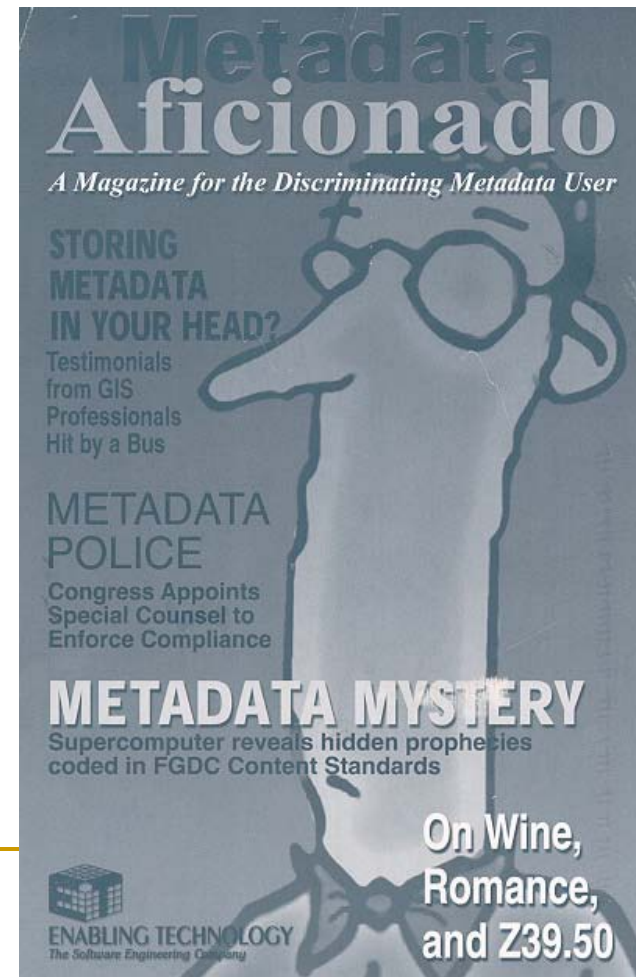


# Metadata Program Overview

## USGS Biological Informatics Metadata Program offers:

- Record creation assistance
- Training
- Metadata creation tools
- Quality Control
- Clearinghouse

*These services are free for USGS and its partner organizations*



# BIP Metadata Program: Key Players

- Viv Hutchison
  - US Geological Survey
  - Seattle, WA
  - Metadata Program Coordinator
  - [vhutchison@usgs.gov](mailto:vhutchison@usgs.gov)
- Giri Palanisamy
  - Oak Ridge National Lab
  - Oak Ridge, TN
  - Clearinghouse technical manager
- Colin Talbert
  - US Geological Survey
  - Fort Collins, CO
  - Quality control
- Alicia Aleman
  - NASA
  - Greenbelt, MD
  - Metadata record creation



Giri



Colin



Alicia

# USGS Biological Informatics Metadata Website

## Metadata

[NBII Home](#) > [Toolkit](#) > [Metadata](#) > [FGDC Metadata](#) > [FGDC Metadata Home](#)

### Metadata

[FGDC Metadata Home](#)

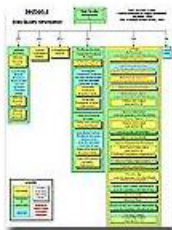
[Discover Data](#)

[Standards](#)

[Tools](#)

[Training](#)

## FGDC Metadata



Metadata is a description of the content, quality, lineage, contact, condition, and other characteristics of data. The description of the data is organized in a standardized format using a common set of terms. Metadata is literally "data about data". Metadata records are similar in concept to library catalog records: details about a book such as title, author, and publisher are recorded in a standard way to ease the search for information.

Biological metadata records work in the same way: information is recorded in a standardized format about a data set (content, quality, condition, and more) for use and analysis. Metadata ultimately makes information about data sets more easily accessible to scientists and researchers.

Metadata is a valuable tool. Metadata records preserve the usefulness of data over time by detailing methods for data collection and data set creation. Metadata greatly minimize duplication of effort in the collection of expensive digital data and foster sharing of digital data resources. Metadata supports local data asset management such as local inventory and data catalogs, and external user communities such as Clearinghouses and websites. It provides adequate guidance for end-use application of data such as detailed lineage and context. Metadata makes it possible for data users to search, retrieve, and evaluate data set information from the NBII's vast network of biological databases by providing standardized descriptions of geospatial and biological data.

The NBII offers an extensive array of metadata resources. Explore the links below and find resources for your organization to use.

- Discover the [NBII Metadata Clearinghouse](#).
- Learn about [Metadata Standards](#).
- Research a variety of [Metadata Creation Tools](#).
- Find [Metadata Training Workshops](#) and resources for trainers.

## Metadata Clearinghouse



The [National Biological Infrastructure \(NBII\) Clearinghouse](#) is an initiative to help you locate, evaluate, and access biological data and information from a distributed network of cooperating data and information sources.

The NBII Clearinghouse includes metadata descriptions of biological databases and information products developed and maintained by USGS scientists, as well as data and information developed and maintained by other NBII participants, including federal, state, and local government agencies; private sector organizations; and other partners around the nation and the world.

- [Search the Clearinghouse](#)
- [Add Your Metadata Record](#)

# USGS Biological Informatics Clearinghouse Dashboard



- Inside this Community -

Select Your Organization  
Current partner selected: Metadata Clearinghouse Principal Node  
Click here to select a different partner.

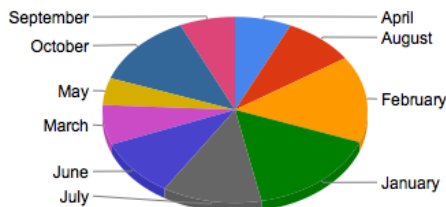
Number of Times Your Records are Accessed

Your Most Accessed Records

Title	Count
Birds of Brasilia	26
Using Incubation Period Length as a Tool to Predict Host Impact	19
Checklist of Amphibian Species and Identification Guide: An Online Guide for the Identification of Amphibians in North America north of Mexico	19
Tennessee State Parks All Taxa Biodiversity Inventory Data	16
Sacramento Refuge Roads	14

Overall Clearinghouse Searches

Overall NBIIHC Performance in 2010



Most Frequent Keywords Occurrences in all records

Keywords	Occurrences
line	767
dlg	430
graph	406

Your Metadata Dashboard

## Welcome to the NBII Clearinghouse Metadata Provider Dashboard

The NBII appreciates your submissions to the Clearinghouse. We are striving to make useful, robust records available to the scientific community. Metadata records are important in that they provide scientists the ability to share detailed information about their datasets. It is important that we, in our data sharing efforts, strive to maintain high quality metadata – both for data discovery and data use purposes. As an aggregator of metadata records, the NBII harvests metadata from a large number of providers. This dashboard has been created to help you, as an organization providing metadata records to NBII, to check on the status and quality of those records. The dashboard allows you to check on your metadata submissions in a variety of ways.

Thank you for your commitment to providing metadata records to the NBII Clearinghouse! We hope you find the Provider Dashboard a useful experience. Check back often, as new content will be added.

Broken Links Found in Your Metadata Records

Printable Version

Title	Link	Date Broken
1989 Land Cover/Land Use on the Upper Mississippi River System	Missing link	Tue Jan 04 17:32:48 EST 2011
1990 land use for part of Muddy Creek basin, Rhode River, MD	Missing link	Tue Jan 04 17:29:08 EST 2011
1995 International Crown Fire Experiment in Northern Canada	Missing link	Tue Jan 04 16:13:04 EST 2011
		Tue Jan 04
		<b>Total broken links: 1909</b>

Your Metadata Records by Title

- A Nationwide Geographical Information System (GIS) Analysis of Mine-Roosting Category 2 (C2) Bat Species
- A Spatial and Genetic Analysis of Cowbird Host Selection
- A watershed approach to ecosystem monitoring in Denali National Park and Preserve, Alaska
- ACE Basin National Estuarine Research Reserve Water Quality Data
- ACE Basin National Estuarine Research Reserve Water Quality Data in 1995
- ACE Basin National Estuarine Research Reserve Water Quality Data in 1996
- Accuracy Assessment Spatial Database for the Voyageurs National Park Vegetation Mapping Project
- Acoustic Doppler Current Profiler Data Sets from Fixed Vessel
- Agate Fossil Beds National Monument, Field Plots Data Base for Vegetation Mappings
- Air quality monitoring protocol - Denali National Park and Preserve
- Alfagenes, An Alfalfa Genome Database Including Cultivar
- An Automated Photo-identification Catalog for Studies of the Life History of the Florida Manatee
- Apalachicola National Estuarine Research Reserve Water Quality Data
- Apalachicola National Estuarine Research Reserve Water Quality Data in 1995

# Metadata Input/Update Tool



[New Basic Metadata](#)

[New Data File Structure](#)

[New Research Site](#)

[New Variable](#)

[New Person](#)

**admin**

- [Biblio](#)
- [Glossary](#)
- [My account](#)
- [Create content](#)
- [Administer](#)
- [Log out](#)

**Metadata Archive**

- [May 2010 \(1\)](#)

**Recent Publications**

- [Fire Regimes of Kipuka Forests in El Malpais National Monument, New Mexico](#)
- [More...](#)




## DIMS Metadata Editor

Submitted by [admin](#) on Mon, 2010-05-24 13:32

You will find here an opportunity to use the Drupal based metadata editor. A number of ir metadata editors based on the Drupal content management system have been deployed (<http://tierra.unm.edu>, <http://gorilla.ites.upr.edu>, <http://lter.iecolab.es:8090/lterSpain>, e features a long list of user oriented features. The following is a reduced list:

- Easy to customize
- Picklists for personnel, research sites, variables and more (auto-completes)
- Controlled vocabularies
- Automated versioning mechanism

# Metadata Input/Update Tool



[New Basic Metadata](#) | [New Data File Structure](#) | [New Research Site](#) | [New Variable](#) | [New Person](#)

[Home](#) » [Create content](#)

## Create Data Set

[Basic Information](#) | [Personnel Involved](#) | [Geo Temporal Info](#) | [Methodologies](#)

**Title:**  
  
The title of the dataset

**Numeric Identifier:**  
  
The Sevilleta ID assigned to the data set - Omit the prefix "SEV", input just the number.

**Short name:**  
  
A shortname is usually like a file name for your dataset container - the name used to refer in short to your dataset. Example: meteo\_2010

**Abstract:**  
  
This is the data set abstract

[▶ Input format](#)

**Dataset EML ID:**  
  
This is a numeric identifier for your data set. It should be unique - that is - you cannot use a previously used identifier. Example : 1012009

**Purpose:**  
  
The purpose of your dataset - the goals sought with the collection of your data.

**Data File Structure:**

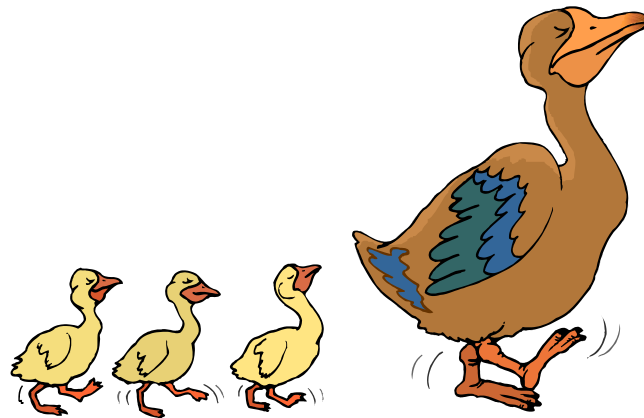
The reference to the information about data file structure details. If the structure has not been entered yet, ignore this for now, and come back later to enter it.

**Bibliography:**



---

# Beyond Clearinghouses



# Uses for metadata depend on how complete the record is...

- Identification
- Discovery
- Evaluation
- Access
- Integration

Completeness Level	Description and Major Elements Added
<b>1: Identification</b>	Minimum content for adequate data set discovery in a general cataloging system or repository (functionally equivalent to LTER DTOC): <ul style="list-style-type: none"> <li>• title</li> <li>• creator</li> <li>• contact</li> <li>• publisher</li> <li>• pubDate</li> <li>• keywords</li> <li>• abstract (recommended)</li> <li>• dataset/distribution (i.e. url for general dataset information)</li> </ul>
<b>2: Discovery</b>	Level 1 content, plus coverage information to support targeted searches <ul style="list-style-type: none"> <li>• geographicCoverage</li> <li>• taxonomicCoverage</li> <li>• temporalCoverage</li> </ul>
<b>3: Evaluation</b>	Level 2 content, plus data set details to enable end-user evaluation of the methodology and data entities: <ul style="list-style-type: none"> <li>• project</li> <li>• methods</li> <li>• entity</li> <li>• attributes (strongly recommended, as possible)</li> <li>• intellectualRights</li> </ul>
<b>4: Access</b>	Level 3 content plus data access details to support computer-assisted data retrieval: <ul style="list-style-type: none"> <li>• access</li> <li>• physical</li> </ul>
<b>5: Integration</b>	Level 4 content plus complete attribute and QA/QC details to support computer-assisted data integration and re-sampling <ul style="list-style-type: none"> <li>• attribute (required)</li> <li>• measurementScale</li> <li>• units</li> <li>• constraint</li> <li>• qualityControl</li> </ul>
<b>6: Semantic Use</b>	Level 5 content plus semantic information (currently under development by SEEK, and may require extension to the EML schema)

---

# Metadata: Critical to Data Integration

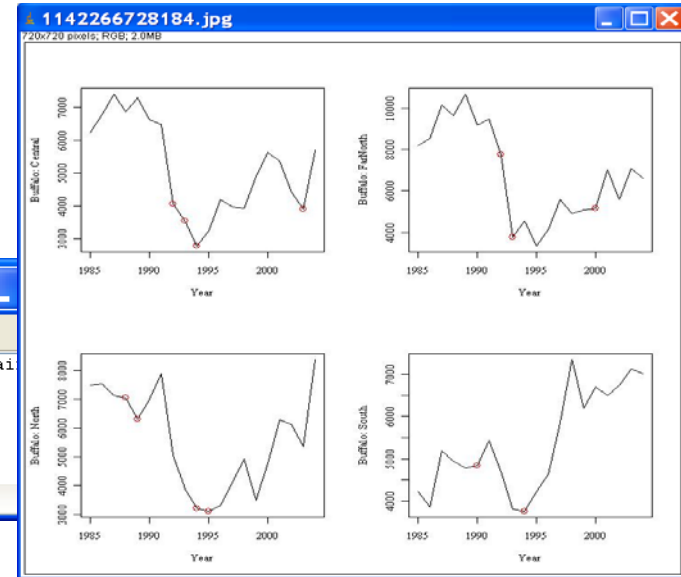
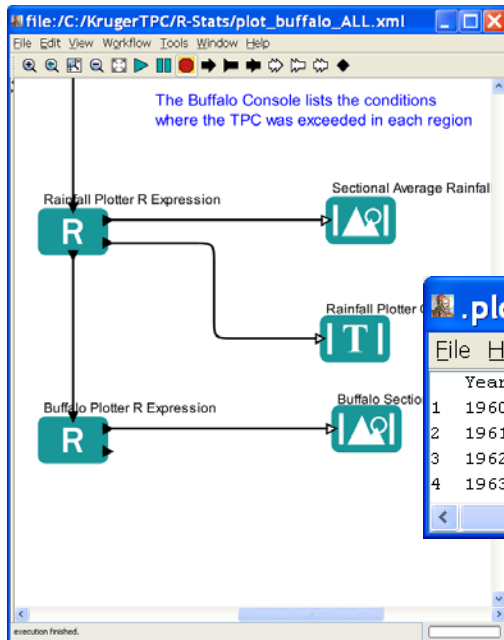
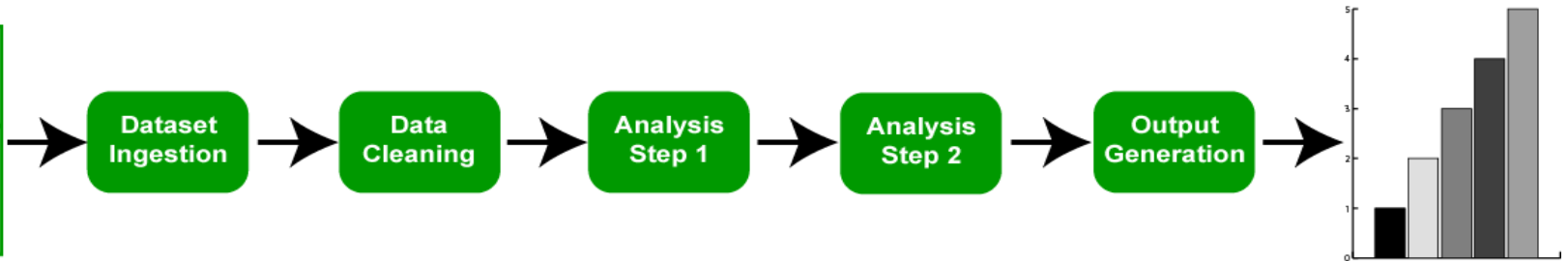
Creating quality metadata: some true payoffs...

- Analysis: Overall, quality metadata makes analysis easier
- Central archives: allow data created in the past to be reused in analysis
- Quality Control:
  - Workflows can be built to find inconsistencies in data by comparing data and metadata (i.e.: if a value exceeds a domain in data, but metadata says it shouldn't)
  - Analysis on two datasets with different units of measure can result in appropriate conversions

# From Clearinghouses to Data Integration

## ▣ Moving beyond discovery:

Raw Data
x, y, z
1, 3, 6
3, 8, 5
1, 9, 0
Metadata
<enl>
<dataset>
<title>Dataset 1</title>
...



---

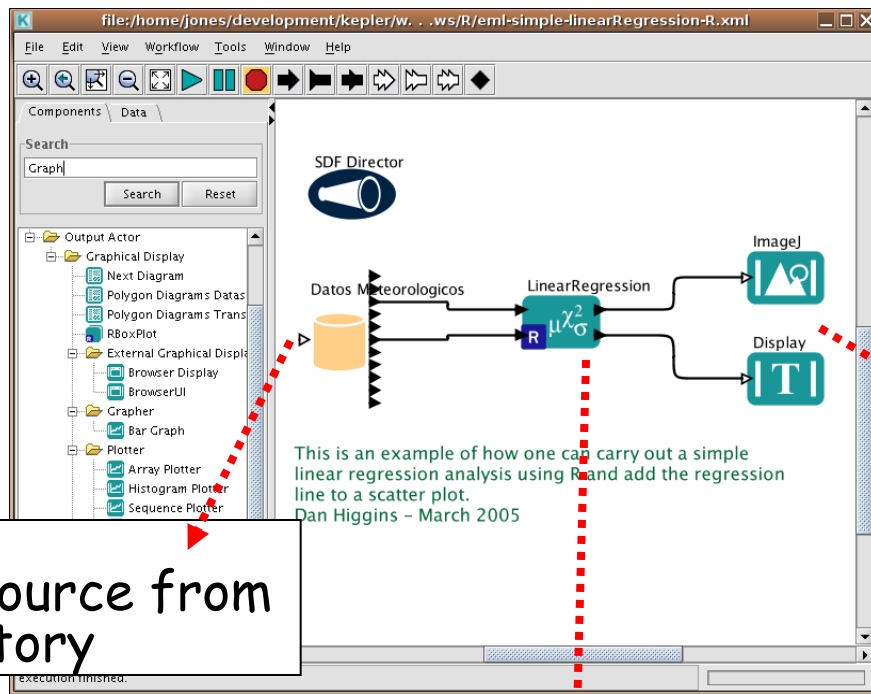
# Kepler scientific workflow system

Kepler is a software application for analyzing and modeling scientific data:

It allows scientists to:

- Create executable models, called "scientific workflows"
- Access scientific data (streaming sensor data, medical and satellite images, simulation output, observational data, etc.)
- Execute complex analyses on this data

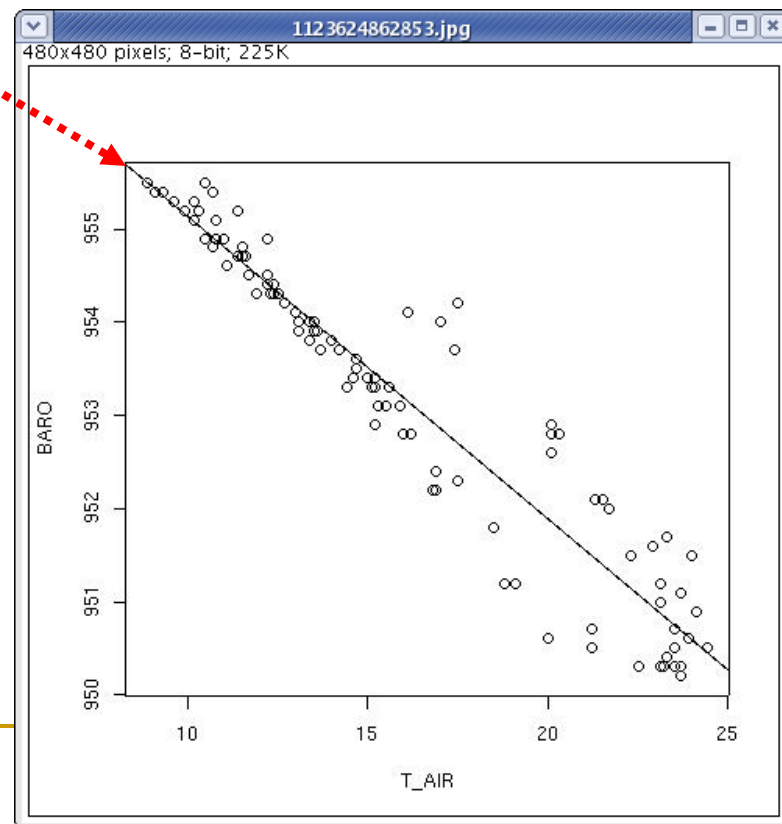
# Kepler scientific workflow system



Data source from repository

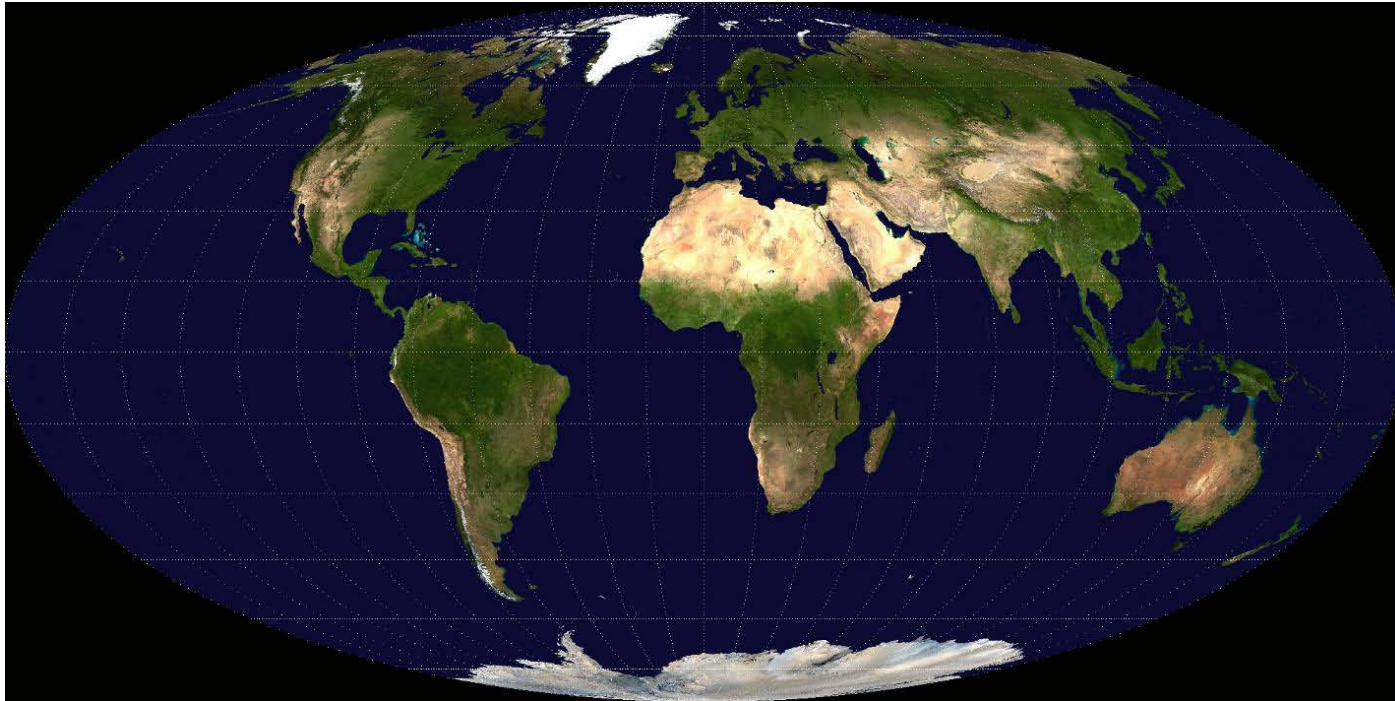
R processing script

```
res <- lm(BARO ~ T_AIR)
res
plot(T_AIR, BARO)
abline(res)
```



---

# DataONE



Building a Global Cyberinfrastructure

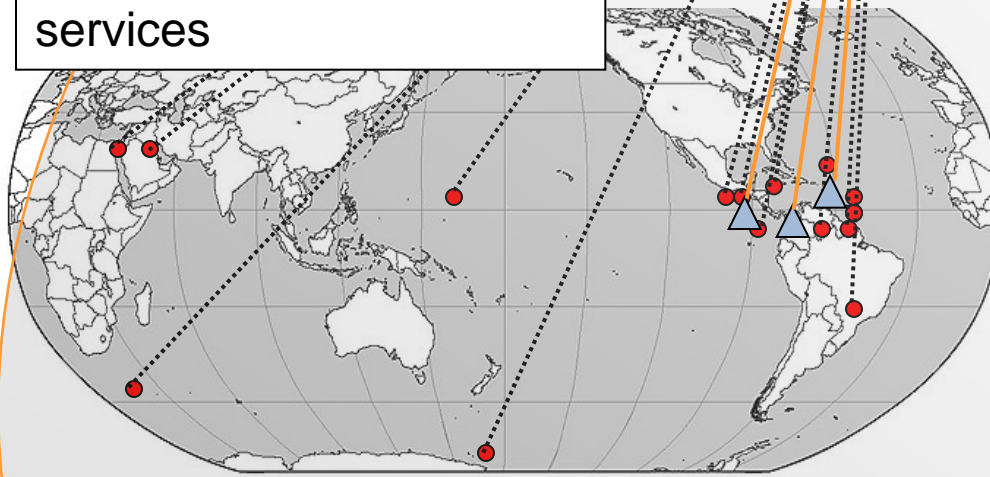
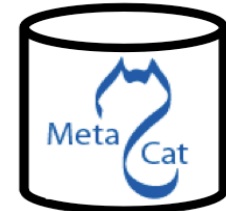
# Distributed framework

## Coordinating Nodes

- retain complete metadata catalog
- subset of all data
- perform basic indexing
- provide network-wide services
- ensure data availability (preservation)
- provide replication services

Flexible, scalable,  
sustainable network

## Investigator 1..N Toolkit

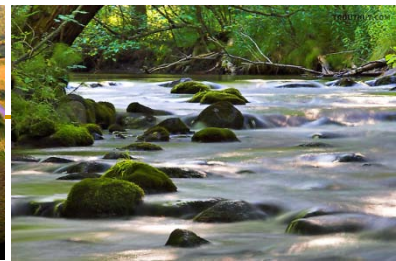
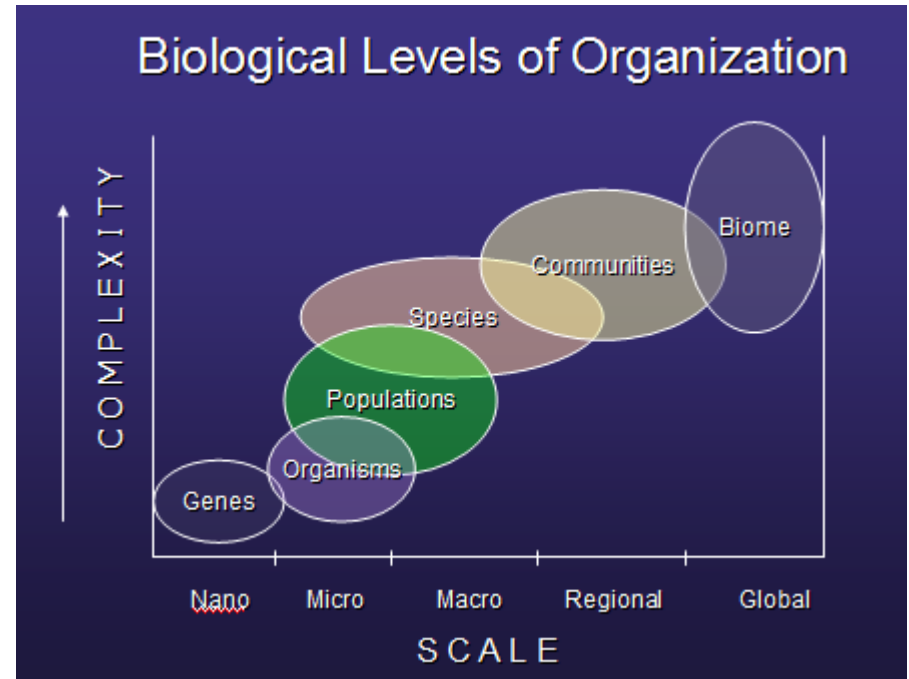


0 5000 Miles

0 5000 Kilometers

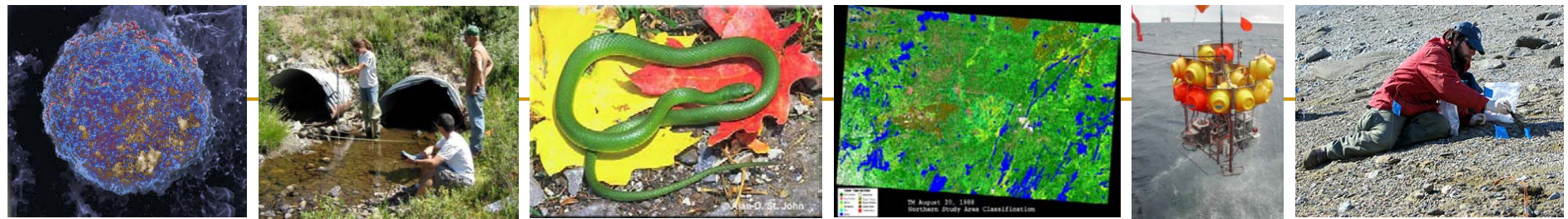
# Data types

- Biological (genes to biomes)
- Environmental
  - Atmospheric
  - Ecological
  - Hydrological
  - Oceanographic










# Data sources

- Research networks and environmental observatories
- Biological specimens
- Individual scientists
- Citizen scientists' data
- Natural resources and conservation data
- Observational data
- Global and continental land cover/land change and biogeochemical data



# Examples of Data Holdings

## Metadata Interoperability Across Data Holdings

Data Archive	Types of Data Managed	Metadata Standard(s)
	Biodiversity, taxonomic, ecological	BDP, DwC, DC, OGIS
	Biogeochemical dynamics, terrestrial ecological Earth observation imagery	DIF, BDP, ECHO
	Ecological, biodiversity, biophysical, social, genomics, and taxonomic	EML
	Avian populations and molecular biology	DwC
	Biological and taxonomic	DC subset
	Biophysical, biodiversity, disturbance, and Earth observation imagery	EML
	Biodiversity, biotic structure, function/process, biogeochemical, climate, and hydrologic	EML



BDP=Biological Data Profile  
DC=Dublin Core

DC subset=Dublin Core subset  
DIF=Directory Interchange Format

DwC=Darwin Core  
ECHO=EOS ClearingHouse  
OGIS=OpenGIS

# Provide One-Stop Shopping for Data

Pilot Catalog: Simple Search Interface  
(searches entire metadata record)

>48,000 Data Set Records

The screenshot shows the DataONE Pilot Catalog search interface. At the top left is the DataONE logo, and at the top right is the text "A Pilot Catalog For Earth Observations". Below this is a navigation bar with "DataONE Metadata Clearinghouse" and a "HELP" link. The main search area has two tabs: "Simple Search" (selected) and "Advanced Search". The search prompt is "Search All Records For". There is a search input field, a "SEARCH" button, and a "Results/Page" dropdown menu set to "10". A hint text below the input field reads: "Hint: boolean operators, wildcards and phrases are allowed. ex: precipitation or (rain\* and \"moisture content\")". Below the search area is a section labeled "Query being built:" with a "Not Editable" status and a "CLEAR QUERY" link.

**NBII Metadata Clearinghouse (39,550)**

**Long Term Ecological Research (LTER) Network (6,897)**

**ORNL Distributed Active Archive Center for Biogeochemical Data (810)**

**Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA) (783)**

**Organization of Biological Field Stations (124)**

**Inter-American Institute for Global Change Research (IAI) (79)**

**MODIS and ASTER Products (LPDAAC) (38)**

**National Phenology Network (USANPN) (29)**

# Enabling tools



**Morpho**  
 Data Management for Ecologists

Morpho is a data management tool for ecologists. The screenshot shows a butterfly image and a table of data. The table has columns for 'Temperature Monitored' and 'Access Control'.

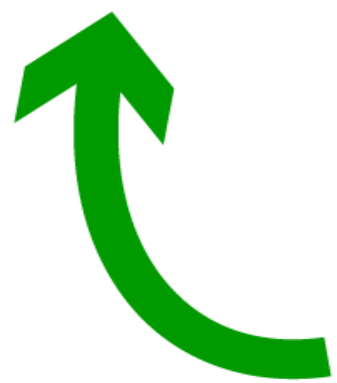
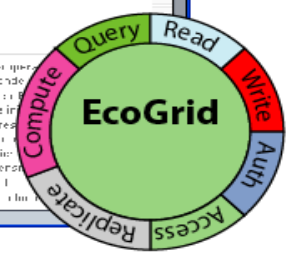
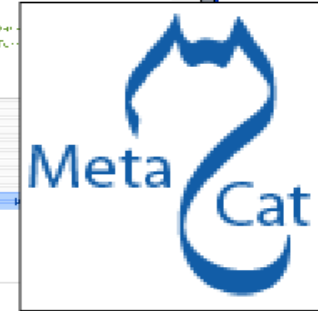
Temperature Monitored	Access Control
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	
1998-01-01 08:40:00 1998-02-01 08:40:00 APTX null null 01	

**Biocomplexity Data Search**  
 Home

Search for data on the KND. You are logged in (Logout). Enter a search phrase (e.g. odonata) to search for data sets by author, title, label, email, or simply access to odonata.

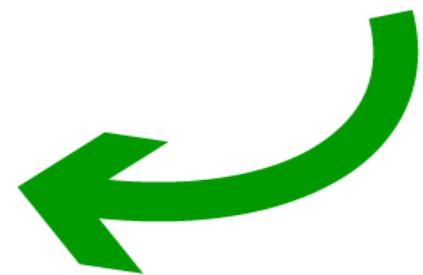
**456 data packages found**

Title	Contact	Organization
Buller's antechinus	Yuan Feng	
Productivity, Diversity and Soil Data from two North American Grasslands	Doe	



**Kepler**

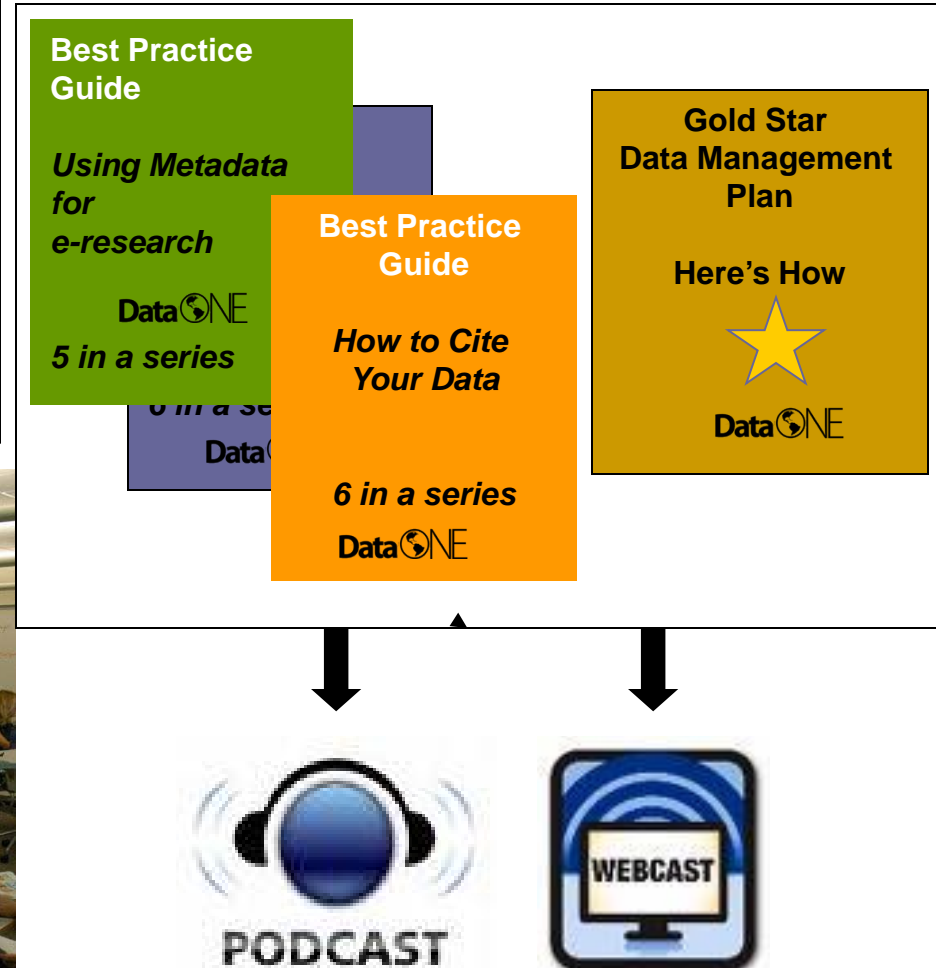
The screenshot shows a workflow in the Kepler environment. It includes components like 'Gene Accession Number', 'Gene Accession Number and Sequences Display', 'Extract Gene Sequence', and 'Sequence Transfer'. The workflow is visualized as a series of connected nodes and arrows.



# Education and training

## Career Long Learning:

- best practice guides
- exemplary data management plans
- podcasts, web-casts
- workshops and seminars
- downloadable curricula



# Building global communities of practice:

## ... creating long-lived CI enterprises

- Broad, active community engagement
  - Involvement of library and science educators engaging new generations of students in best practices
  - Existing outreach and education programs
- Transparent, participatory governance
- Adoption/creation of innovative and sustainable business and organizational models



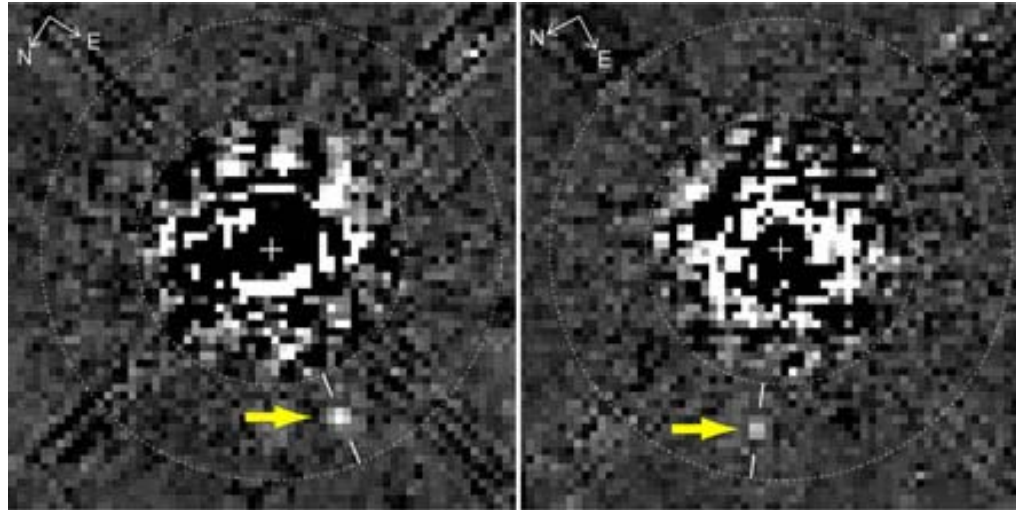
# ... engaging diverse partners

- ❑ Libraries & digital libraries
- ❑ Academic institutions
- ❑ Research networks
- ❑ NSF- and government-funded synthesis & supercomputer centers/networks
- ❑ Governmental organizations
- ❑ International organizations
- ❑ Data and metadata archives
- ❑ Professional societies
- ❑ NGOs
- ❑ Commercial sector



# “Planet hidden in Hubble archives” *Science News* (Feb. 27, 2009)

- A new image processing technique reveals something not before seen in this Hubble Space Telescope image taken 11 years ago: A faint planet (arrows), the outermost of three discovered with ground-based telescopes last year around the young star HR 8799. D. Lafrenière et al., *Astrophysical Journal Letters*



“The first thing it tells you is how valuable maintaining long-term archives can be. Here is a major discovery that’s been lurking in the data for about 10 years!” comments Matt Mountain, director of the Space Telescope Science Institute in Baltimore, which operates Hubble.

“The second thing its tells you is having a well calibrated archive is necessary but not sufficient to make breakthroughs — it also takes a very innovative group of people to develop very smart extraction routines that can get rid of all the artifacts to reveal the planet hidden under all that telescope and detector structure.”

---

# DataONE Leadership Team

## Leadership Team:

Suzie Allard – *Univ. of Tenn.*

John Cobb – *Oak Ridge Nat'l Lab*

Bob Cook – *Oak Ridge Nat'l Lab*

Patricia Cruse – *Univ. CA Digital Curation  
Center*

Mike Frame – *USGS*

Stephanie Hampton – *Univ. CA Santa Barbara*

Viv Hutchison - *USGS*

Matt Jones – *Univ. CA Santa Barbara*

Steve Kelling – *Cornell Univ.*

John Kunze – *Univ CA Digital Curation Center*

Carol Tenopir – *Univ. of Tenn.*

Bruce Wilson – *Joint Oak Ridge Nat'l Lab  
and Univ. of Tenn.*

Todd Vision – *Duke Univ.*

Bertram Ludascher – *Univ. of CA - Davis*

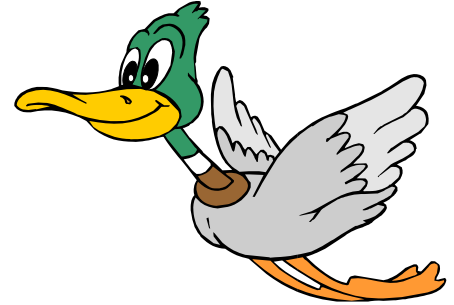


---

# Conclusions

**In this session we reviewed:**

- **Definition of metadata**
- **Value of metadata**
- **Working with the FGDC metadata standard**
- **Clearinghouses**
- **Implementation tips**
- **Beyond Metadata**



# Thank you!

## Questions? Comments?

Viv Hutchison

USGS – Denver, CO

Contact: [vhutchison@usgs.gov](mailto:vhutchison@usgs.gov) or  
303.202.4227

