Enabling structured and unstructured searches of your SDI components

David Danko
ISO Metadata Project Leader
Senior Consultant – GIS Standards, ESRI
Data about Data (official definition)
Information that describes a resource (data) so it can be understood
Ensures we find and use the right data for the right purpose
And used correctly
Geographic Metadata

- Data about data
  - Geographic area covered
  - Currency
  - Rules of use
  - Positional accuracy
  - Data dictionary
  - Means of encoding
  - Datum
  - Map projection

- A Key interoperability technology:
  - Producers explain their product
  - Users learn about products
Would you consume anything without knowing what it is?
METADATA USES
UNDERSTANDING THE RIGHT DATA FOR THE RIGHT PURPOSE

❖ Locate
♦ Find
♦ Discover
♦ Structured searches vs. unstructured

❖ Extract
♦ Order
♦ Download

❖ Evaluate
♦ Restrictions
♦ Quality
♦ Reputation

❖ Employ
♦ Define,
♦ Apply,
♦ Use,
♦ & Understand geographic knowledge
ISO 19115 RECOMMENDED CORE

This list contains metadata elements answering the following questions: “Does a dataset on a specific topic exist (‘what’)?”; “For a specific place (‘where’)?”; “For a specific date or period (‘when’)?” and “A point of contact to learn more about or order the dataset (‘who’)?”.

- Dataset title
- Dataset reference date
- Dataset responsible party
- Geographic location of the dataset (by four coordinates or by geographic identifiers)
- Dataset language
- Dataset character set
- Dataset topic category
- Spatial Resolution
- Abstract describing the dataset
- Distribution format

- Additional extent information (vertical and temporal)
- Spatial representation type
- Reference system
- Lineage statement
- On-line resource
- Metadata file identifier
- Metadata standard name
- Metadata standard version
- Metadata language
- Metadata character set
- Metadata point of contact
- Metadata date stamp

✔ Mandatory

Automated

Copyright © 2008 ESRI. All rights reserved.
1. farming
2. biota
3. boundaries
4. climatology/Meteorology
5. economy
6. elevation
7. environment
8. geoscientificInformation
9. health
10. imagery/BaseMapsEarth Cover
11. intelligence/Military
12. inlandWaters
13. location
14. oceans
15. planning/Cadastre
16. society
17. structure
18. transportation
19. utilities/Communication
<table>
<thead>
<tr>
<th>Topic Category Code</th>
<th>Topic Cat Cd</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>farming</td>
<td>001</td>
<td>rearing of animals and/or cultivation of plants. Examples: agriculture, irrigation, aquaculture, plantations, herding, pests and diseases affecting crops and livestock.</td>
</tr>
<tr>
<td>biota</td>
<td>002</td>
<td>flora and/or fauna in natural environment. Examples: wildlife, vegetation, biological sciences, ecology, wilderness, sealife, wetlands, habitat.</td>
</tr>
<tr>
<td>boundaries</td>
<td>003</td>
<td>legal land descriptions. Examples: political and administrative boundaries.</td>
</tr>
<tr>
<td>climatology, meteorology, atmosphere</td>
<td>004</td>
<td>processes and phenomena of the atmosphere. Examples: cloud cover, weather, climate, atmospheric conditions, climate change, precipitation.</td>
</tr>
<tr>
<td>economy</td>
<td>005</td>
<td>economic activities, conditions and employment. Examples: production, labour, revenue, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, exploration and exploitation of resources such as minerals, oil and gas.</td>
</tr>
<tr>
<td>elevation</td>
<td>006</td>
<td>height above or below sea level. Examples: altitude, bathymetry, digital elevation models, slope, derived products.</td>
</tr>
<tr>
<td>environment</td>
<td>007</td>
<td>environmental resources, protection and conservation. Examples: environmental pollution, waste storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape.</td>
</tr>
<tr>
<td>geoscientific information</td>
<td>008</td>
<td>information pertaining to earth sciences. Examples: geophysical features and processes, geology, minerals, sciences.</td>
</tr>
</tbody>
</table>
GIS Portal Enables ...

- Faster discovery
- Direct access and use
- Collaboration for new data collection
- Improvement of data quality and coverage

...Organizes a Spatial Data Infrastructure
Searching Metadata

- Defines a common interface that enables diverse but conformant applications to perform discovery, browse and query operations against distributed and potentially heterogeneous catalog servers.

- Three Parts:
  - Discovery Services: allow a client to locate metadata that describes data.
  - Access Services: provide the client with methods to request services on the data. Direct Access provides the client with a handle which, when used by the client, provides data to the client.
  - Management Services: defines methods for a client to change the metadata held by a catalog.

- There are both tightly coupled (like CORBA) and loosely coupled (HTTP/XML – CS-Web) profiles.

- Some bindings defined:
  - Z39.50,
  - CS-W: ebRIM, ISO Metadata, OGC Core
Understanding OGC Catalog Services

- **Resource**
  - Data
  - Service (WMS, WFS, WCS, ...)

- **Metadata**
  - FGDC ISO 19115, 19119

- **Client**

- **Service Requester**

- **Service Provider**

- **Service**

- **Serves**

- **Uses**

- **Mapped to**

- **Catalog**

- **Based On**

- **Specific Data Model**

- **Registry Information Model**
  - eb-RIM
  - ISO
  - FGDC
  - Dublin Core

Copyright © 2008 ESRI. All rights reserved.
Structured vs. Unstructured

Structured search
- Metadata is produced about a resource
- Menu/template provided to search the metadata
- Search can contain multiple variables

Unstructured search
- Metadata derived from a resource
- If metadata is not apparent in the resource it is not produced
- Simple searches
### Advanced Search

**Find results**
- with all of the words
- with the exact phrase
- with at least one of the words
- without the words

**Language**
- Return pages written in
- any language

**Region**
- Search pages located in:
- any region

**File Format**
- Only return results of the file format
- any format

**Date**
- Return web pages first seen in the
- anytime

**Occurrences**
- Return results where my terms occur
- anywhere in the page

**Domain**
- Only return results from the site or domain
- e.g. google.com, .org

**Usage Rights**
- Return results that are free to use share or modify
- More info

**SafeSearch**
- No filtering
- Filter using SafeSearch

---

### Page-Specific Search

- Find pages similar to the page

- Find pages that link to the page

---

### Topic-Specific Searches

- Search the full text of books
- Search public source code
- Search scholarly papers
- Search historical news
- Search for all things Mac
- Search web pages about the BSD operating system
- Search all penguin-friendly pages
- Search Microsoft-related pages
- Search a specific school's website
Only formats that can be determined by .suffix
Identifies websites whose owners have indicated that they carry a Creative Commons license.
ISO 19115 Recommended Core

Dataset title
Dataset reference date
Dataset responsible party
Geographic location of the dataset (by four coordinates or by geographic identifiers)
Dataset language
Dataset character set
Dataset topic category
Spatial Resolution
Abstract describing the dataset
Distribution format
Additional extent information (vertical and temporal)
Spatial representation type
Reference system
Lineage statement
On-line resource
Metadata file identifier
Metadata standard name
Metadata standard version
Metadata language
Metadata character set
Metadata point of contact
Metadata date stamp
ISO 19115 Content

- Metadata Entity Set Information
- Identification Information
- Constraints
- Data Quality Information
- Maintenance Information
- Spatial Representation Information
- Reference System Information
- Content Information
- Portrayal Catalogue Information
- Distribution Information
- Application Schema Information

Copyright © 2008 ESRI. All rights reserved.
Without metadata both structured and unstructured searches would be impossible.

Robust metadata supports robust searches.

Metadata lets us better understand the resources we discover and to use them properly.