OpenGeoportal Metadata

Marc McGee
Harvard Library
OGP Metadata Working Group
Metadata Standards

- MARC21 – Library cataloging standard
- MODS
- Dublin Core / DCLite4G
- Many other possibilities ...
Critical Elements for Discovery and Identification

- Title
- Spatial extent – bounding box coordinates
- Originator(s)
- Publisher
- Date
- Keywords – theme & place
- Data type
- Access restriction
- Host repository
- Online location
OGP Solr Schema

• XML-based, layer-level description

• Indexes the critical discovery elements

• Stores original metadata for display/delivery

https://code.google.com/p/opengeoportal/wiki/Solr
Egyptian Cities from The Multipurpose Africover Database for the Environmental Resources produced by the Food and Agriculture Organization of the United Nations (FAO). Scale of the dataset: 1:200,000.

The towns and major towns shapefiles were developed from data from the National Imagery and Mapping Agency (NIMA).
**Egypt Cities**

**Metadata:**
- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Metadata Reference Information

**Identification Information:**

**Citation:**

```
Originator: Food and Agriculture Organization of the United Nations
Originator: Di Gregorio, Antonio
Originator: Egypt. Ministry of Agriculture. Soil and Water Research Institute
Originator: El Mowelhi, Nabil
Publication Date: 20020404
Title: Egypt Cities
Edition: 1st
Geospatial Data Presentation Form: vector digital data
```

**Publication Information:**

```
Publication Place: Cambridge, Massachusetts
Publisher: Harvard Map Collection, Harvard College Library
```

**Online Linkage:**

```
http://hgl.harvard.edu:8080/HGL/isp/HGL.jsp?action=VColl&VCollName=AFRICOVER_EG_CITY
```

**Description:**

```
Egypt cities from The Multipurpose Africover Database for the Environmental Resources produced by the Food and Agriculture Organization of the United Nations (FAO). Scale of the dataset: 1:200,000. The towns and major towns shapefiles were developed from data from the National Imagery and
```
Shared Conventions for Content

• **Title** – Feature type, Location, Date (not filename)

• **Originators** – ESRI, E.S.R.I., Esri, ESRI Inc. Environmental Systems Research Institute (Redlands, Calif.)

• **Theme/place keywords** – ISO 19115 theme keys, GNIS/GNS gazetteers

• **Publishers** – U.S. Census Bureau, Census Bureau ...

• **Data type** – vector, raster, tabular, map (georeferenced, digital, paper), remote-sensing image ...

• **Coordinates** – bounding box in decimal degrees
Critical Elements for Dataset Use

- Abstract / purpose of dataset
- Terms of use
- Spatial references - coord system / projection
- Spatial data organization – vector type / features; raster type / resolution
- Source information – citation, scale, date, contribution
- Data quality information (if available) – accuracy and completeness
- Process steps
- Attribute definitions
Details for CH2000_A_A

**Description**

**Attributes**

<table>
<thead>
<tr>
<th>CNAME</th>
<th>GBCNTY</th>
<th>GBPROV</th>
<th>GBCITY</th>
<th>CH2000_A_A.AREA</th>
</tr>
</thead>
</table>

**ENAME**

- **A101001**
  - **Definition:** Total N of Households by County
  - **Definition Source:** China Data Center

- **A101002**
  - **Definition:** Total N of Family Households by County
  - **Definition Source:** China Data Center

- **A101003**
  - **Definition:** Total N of College Households by County
  - **Definition Source:** China Data Center

<table>
<thead>
<tr>
<th>ENAME</th>
<th>A101001</th>
<th>A101002</th>
<th>A101003</th>
<th>A101004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehui Shi</td>
<td>239030</td>
<td>237531</td>
<td>1499</td>
<td>878349</td>
</tr>
<tr>
<td>Dejiang Xian</td>
<td>106169</td>
<td>106062</td>
<td>107</td>
<td>388639</td>
</tr>
<tr>
<td>Delingha Shi</td>
<td>26681</td>
<td>24249</td>
<td>2432</td>
<td>83585</td>
</tr>
<tr>
<td>Dengfeng Shi</td>
<td>151689</td>
<td>147333</td>
<td>4356</td>
<td>609085</td>
</tr>
<tr>
<td>Dengkou Xian</td>
<td>35694</td>
<td>35317</td>
<td>377</td>
<td>117561</td>
</tr>
<tr>
<td>Dengta Shi</td>
<td>157322</td>
<td>156225</td>
<td>1097</td>
<td>502149</td>
</tr>
</tbody>
</table>

**A101004**

- **Definition:** Total Population by County
- **Definition Source:** China Data Center
Metadata Creation Workflow

• Gather existing metadata / documentation / code books related to dataset

• Create/apply metadata templates esp. for series-type data or for institutionally appropriate information

• Layer-level description in XML using local standards
  – e.g. title, abstract, access, keyword thesauri, local processing steps, etc.
Metadata Creation Workflow pt. II

• Quality control / quality assurance / XML validation

• Load metadata into OGP
Metadata Authoring Tools

• Combination of ArcCatalog, Extensions, & text/xml editors

• Overview of OGP contributors’ md practices: http://tiny.cc/4asv7w [google doc]

• OGP Metadata Toolkit
  – Subset of critical elements
  – Guided metadata creation
  – Search and re-use existing metadata
  – Transform between metadata standards
Metadata Working Group (MWG)

• Membership, Governance, Notes from National Summit
  http://opengeoportal.org/working-groups/metadata/

• Communication on the general OGP listserv
  http://opengeoportal.org/listservs/

• Voting members to meet quarterly virtually– membership is voluntary & open
MWG Initiatives

• Best practices and standards documentation
  – Spatial metadata, FGDC, ISO, MARC, MODS, guide for vendors/data providers
• Thesaurus & controlled vocabularies
• Metadata creation coordination:
  Tracking Sheet: http://tiny.cc/fr4q6w
  Tracking Guide: http://tiny.cc/572j6w
• Metadata registry : http://tiny.cc/r44q6w
• MWG Action Item tracking : http://tiny.cc/p73u6w
OGP Metadata Resources

• Metadata Creation Guide
  http://opengeoportal.org/working-groups/metadata/metadata-creation-guide/

• Metadata Creation Resources
  http://opengeoportal.org/working-groups/metadata/metadata-resources/

• OGP listserv
Future enhancements with User-generated data and metadata

- User ratings + use statistics + co-occurrence of downloads = data recommendations

- User comments
  How did you use the data? What were the limitations/results?

- Data consumers into data creators/catalogers
  Metadata toolkit for data creators
Questions? Discussion ...

Marc McGee
Geospatial Resources Cataloger
Harvard Library
OGP Metadata Working Group
mmcgee@fas.harvard.edu