DMS: Biological Data

Environmental Disasters Data Management Workshop (EDDM)
National Conservation Training Center (NCTC) USFWS, Shepherdstown, WV
16-17 September 2014

Felimon Gayanilo
Systems Architect
Harte Research Institute, Texas A&M University Corpus Christi
(email: fgayanilo@tamu.edu)
Overview

“…to quickly reach a science-based consensus about the defining characteristics and regulating processes of an ecosystem to address environmental disaster due to natural or anthropogenic causes using best available technology and data.”
Data Life Cycle

- Plan
- Collect/Generate
- Process/Analyze
- Archive
- Discovery/Reuse
Data Life Cycle

Data Aggregation/Collection/Archive Centers: Biological Data

**Type 1: Desktop/Stand-alone**
Data is stored on off-line services

**Type 2: Short-term funded projects**
Objective-based; data encoded offline and sometime served over the Internet

**Type 3: Institutional and Long-term initiatives**
BCO-DMO; GRIIDC; DataONE; eBIRD
Academic/Research Institutes; LTER;
IEDA; ACADIS; fishbase.org; ecosystemresearch.org

**Type 4: Federal, regional and state programs**
NOAA’s NODC/NGDC/NCDC/Fisheries/ESI/DIVERS;
USDA; NIH NCBI; IOOS; GCOOS TOAST; Dept. of Ecology, WA; [http://www.ecy.wa.gov/database.html](http://www.ecy.wa.gov/database.html)

**Type 5: Multi-national programs**
GBIF/Pangaea/CGIAR/UN Data/ World Bank
Environmental Data
## Processes: Most Common

### Planning
- CDL
- DMPTools
- Online || Templates

### Collection/Generation
- Objective-based (Forms)

### Processing
- Objective-based
- Raw data
- Processed data
- Metadata (ISO/EML/CSDGM)
- Publications

### Archiving/Storing
- Dataset

### Distribution/Discovery
- Desktop
  - Institutional (DataONE)
  - Programs (BCO-DMO, GRIIDC)
  - Federal (NOAA, USDA, EPA, etc.)
  - Multi-national (fishabse.org; GBIF; WB; FAO)

### Objective
- based

### Raw data
- Processed
data

### Metadata
- (ISO/EML/CSDGM)

### Publications

<table>
<thead>
<tr>
<th>Planning</th>
<th>Collection/Generation</th>
<th>Processing</th>
<th>Archiving/Storing</th>
<th>Distribution/Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDL</td>
<td>DMPTools</td>
<td>Online</td>
<td>Objective-based</td>
<td>Objective-based</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Templates</td>
<td>(Forms)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Raw data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Processed data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Metadata (ISO/EML)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CSDGM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Publications</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dataset</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Desktop</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Institutional</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(DataONE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Programs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(BCO-DMO, GRIIDC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Federal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(NOAA, USDA, EPA, etc.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Multi-national</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(fishabse.org; GBIF; WB; FAO)</td>
<td></td>
</tr>
</tbody>
</table>
Common DMS Elements

User account: Data Provider → Data Submission

Data Review

Data Posting

Website (Search Engine)

User account: Download

LDAP/CAS
InCommon
OpenID

{URL} HTTP; FTP; SFTP; GridFTP

Metadata Review

DOI

CSW || WAF || ERDDAP

ISO/EML/CSDGM

Feedback

LDAP/CAS
InCommon
OpenID
Prevailing Issues

- Insufficient information to establish data provenance (metadata) and data review (quality control)
- Failed to establish a common standard (collection, vocabulary, ontology and structure) throughout inhibiting the re-use/re-purposing of the datasets
- Insufficient interoperability and network capabilities
- Temporal and spatial limited (highly heterogeneous; goal setting)
DMS: Biological Data

Thank you!

Felimon Gayanilo
Systems Architect
Harte Research Institute, Texas A&M University Corpus Christi
(email: fgayanilo@tamu.edu)
ISSUES
- Interoperability
- Archival practice and accountability
- Insufficient feedback