ID | ASR_Number | Sample_Number | QC_Code | Analysis_Request_No | External_Sample_Number | Start_Date
---|------------|---------------|---------|---------------------|------------------------|-----------------------
1  | 1383       | 892           | _       | 1                   | _                      | 08-Aug-2002           
2  | 1383       | 902           | _       | 1                   | _                      | 08-Aug-2002           
3  | 1383       | 912           | _       | 1                   | _                      | 08-Aug-2002           

<table>
<thead>
<tr>
<th>Site#</th>
<th>Date</th>
<th>H20 Temperature</th>
<th>Conductance</th>
<th>Turbidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>KRS-031</td>
<td>05-Sep-2001</td>
<td>23.7</td>
<td>380</td>
<td>19</td>
</tr>
<tr>
<td>KRS-030</td>
<td>28-Aug-2001</td>
<td>21.1</td>
<td>1172</td>
<td>25</td>
</tr>
<tr>
<td>KRS-029</td>
<td>22-Aug-2001</td>
<td>25.5</td>
<td>395</td>
<td>14</td>
</tr>
</tbody>
</table>

Debbie Baker, 2005 Kansas Biological Survey lunch talk
<table>
<thead>
<tr>
<th>description</th>
<th>Stream chemistry in Kansas</th>
</tr>
</thead>
<tbody>
<tr>
<td>creator</td>
<td>Don Huggins</td>
</tr>
<tr>
<td>contact</td>
<td>Debbie Baker</td>
</tr>
<tr>
<td>keywords</td>
<td>Kansas, wadeable streams, nutrients…</td>
</tr>
<tr>
<td>dates</td>
<td>1999 - 2002</td>
</tr>
<tr>
<td>methods</td>
<td>Horiba used to measure parameters…</td>
</tr>
<tr>
<td>fields</td>
<td>H2O temperature = water temperature</td>
</tr>
<tr>
<td>units</td>
<td>degrees Celsius</td>
</tr>
</tbody>
</table>

Debbie Baker, CPCB, KBS
Life without metadata
Life with metadata

• Information describing data content, context, quality, structure and accessibility (Michener 2000).
Topics

• Importance of Metadata
• Elements of Metadata
• Results of Metadata
• Morpho – Metadata Software
• Metacat – Metadata Cataloging Software
• KBS Task Force on Databases
Second KNB Data Management Workshop
(Knowledge Network for Biocomplexity Project)

2 - 4 February 2005

LTER Network Office
University of New Mexico, Albuquerque
Topics

• Importance of Metadata
• Elements of Metadata
• Results of Metadata
• Morpho – Metadata Software
• Metacat – Metadata Cataloging Software
• KBS Task Force on Databases
Information Entropy over Time

entropy: a process of degradation or running down or a trend to disorder
– Merriam-Webster

Time of publication

Specific details

General details

Retirement or career change

Accident

Death

after Michener et al., 1997
# LTER Tiered Trajectory for Metadata

<table>
<thead>
<tr>
<th>Tiered Trajectory</th>
<th>Discovery</th>
<th>Access</th>
<th>Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tier 1</strong></td>
<td>Unstructured, online site catalog with minimal metadata</td>
<td>Establish data access policy</td>
<td>Unstructured, machine-readable metadata and data</td>
</tr>
<tr>
<td></td>
<td>Data discovery through manual searches</td>
<td>Data and metadata access requires human intervention</td>
<td>Data use requires human intervention</td>
</tr>
<tr>
<td><strong>Tier 2</strong></td>
<td>Online, enhanced metadata with consistent internal structure</td>
<td>Automated access to data</td>
<td>Structured, comprehensive metadata and data</td>
</tr>
<tr>
<td></td>
<td>Data discovery through machine search</td>
<td>Access to site data and metadata does not require human intervention</td>
<td>Data use does not require human intervention</td>
</tr>
<tr>
<td><strong>Tier 3</strong></td>
<td>Discovery enabling metadata structured in EML</td>
<td>Access enabling metadata structured in EML</td>
<td>Complete validated EML</td>
</tr>
<tr>
<td></td>
<td>Data discovery integrated across network</td>
<td>Data access is integrated across network</td>
<td>Data analysis is integrated across the network</td>
</tr>
<tr>
<td><strong>Future Outcome</strong></td>
<td>Semantic-based discovery through machine-based searches</td>
<td>Data access through a knowledge-based query process</td>
<td>Semi-automated knowledge extraction</td>
</tr>
</tbody>
</table>
Topics

• Importance of Metadata
• **Elements of Metadata**
• Results of Metadata
• Morpho – Metadata Software
• Metacat – Metadata Cataloging Software
• KBS Task Force on Databases
Best Practices – Metadata completeness

- Identification
- Discovery
- Evaluation
- Access
- Integration
- Semantic Use

<table>
<thead>
<tr>
<th>Completeness Level</th>
<th>Description and Major Elements Added</th>
</tr>
</thead>
</table>
| 1: Identification  | Minimum content for adequate data set discovery in a general cataloging system or repository (functionally equivalent to LTER DTOC):  
  - title  
  - creator  
  - contact  
  - publisher  
  - pubDate  
  - keywords  
  - abstract (recommended)  
  - dataset/distribution (i.e., url for general dataset information) |
| 2: Discovery       | Level 1 content, plus coverage information to support targeted searches  
  - geographicCoverage  
  - taxonomicCoverage  
  - temporalCoverage |
| 3: Evaluation      | Level 2 content, plus data set details to enable end-user evaluation of the methodology and data entities:  
  - project  
  - methods  
  - entity  
  - attributes (strongly recommended, as possible)  
  - intellectualRights |
| 4: Access          | Level 3 content plus data access details to support computer-assisted data retrieval:  
  - access  
  - physical |
| 5: Integration     | Level 4 content plus complete attribute and QA/QC details to support computer-assisted data integration and re-sampling:  
  - attributes (required)  
  - measurementScale  
  - units  
  - constraint  
  - qualityControl |
| 6: Semantic Use    | Level 5 content plus semantic information (currently under development by SEEK, and may require extension to the EML schema) |
Level 1 - Identification

- **Description** – Minimum content for adequate dataset discovery

- **Major Elements Added**:
  - Title
  - Creator
  - Contact
  - Publisher
  - Publication Date
  - Keywords
  - Abstract
  - Dataset/distribution (i.e. URL for dataset information)
Level 2 - Discovery

• **Description** – Level 1 content, plus coverage information to support targeted searches

• **Major Elements Added**:
  – Geographic Coverage
  – Taxonomic Coverage
  – Temporal Coverage
Level 3 - Evaluation

• **Description** – Level 2 content, plus data set details to enable end-user evaluation of the methodology and data entities

• **Major Elements Added**:  
  – Intellectual Rights  
  – Project  
  – Methods  
  – Data Table/Entity Group  
  – Data Table/Attributes
Level 4 - Access

- **Description** – Level 3 content plus data access details to support automated data retrieval
- **Major Elements Added**:  
  - Access  
  - Physical
Level 5 - Integration

• **Description** – Level 4 content plus requires that all aspects of the data package be fully described such as complete attribute and quality control details. Supports computer-mediated access and processing of data.

• **Major Elements Added**:
  – Attribute List (full descriptions)
  – Measurement Scale
  – Units
  – Constraint
  – Quality Control
Level 6 - Semantic

• **Description** – Level 5 content plus semantic information (*make everyone use the same verbage*). Currently under development by SEEK (*KUNHM!*!) and may require extension to the EML schema.

Baker – what the ?#)@! are you talking about? What is EML?
Ecological Metadata Language

• A method for formalizing and standardizing the set of concepts that are essential for describing ecological data, as well as the format for recording this information.
• Address lack of dataset documentation.
• Provide structure to traditionally unstructured information.

You really don’t need to know EML. Just use the software we’ll discuss later.
EML Continued
Related metadata standards

- **Dublin Core Element Set**
  - Corresponds roughly to eml-resource

- **Content Standard for Digital Geospatial Metadata (CSDGM)**
  - Federal Geographic Data Committee (FGDC)
  - Corresponds to eml-spatialRaster, eml-spatialVector, eml-spatialReference
  - Overlaps in other modules (eml-resource)

- **Biological Data Profile (BDP) of the CSDGM**
  - Biological Data Working Group of the FGDC
  - Shares structure for taxonomicCoverage, geologicAge, and ascii table structures

- **ISO 19115 Geographic information: Metadata**
  - Incorporated in the eml-spatial* modules
  - Eml-party derived from ISO 19115

- **Darwin Core**
  - Partially overlaps with eml-coverage

- **Geography Markup Language (GML)**
Topics

• Importance of Metadata
• Elements of Metadata
• Results of Metadata
• Morpho – Metadata Software
• Metacat – Metadata Cataloging Software
• KBS Task Force on Databases
Results

• A self-explanatory dataset or database
Results

- A standard, searchable information format.
  - Morpho: a software program into which you enter metadata
  - Metacat: a catalog of metadata files (created in Morpho) that you can search over the web.
Topics

• Importance of Metadata
• Elements of Metadata
• Results of Metadata
• **Morpho – Metadata Software**
• Metacat – Metadata Cataloging Software
• KBS Metadata Committee
Morpho

http://knb.ecoinformatics.org/morphoportal.jsp
Welcome to Morpho
Adding a title

• Begin by adding the title and an abstract.

• Click next to add Keywords.
Adding geographic metadata

- A new window will open.
- You can choose from a predefined list or create a bounding box using lat long values.
Methods and sampling metadata

Enter method step description. Method steps describe a single step in the implementation of a methodology for an experiment.

<table>
<thead>
<tr>
<th>Method Step Title</th>
<th>Method Step Description</th>
<th>Instrumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locating the NPP Sampling Quadrats</td>
<td>Each core site contains five rodent trapp...</td>
<td>Frames - The frame is constructed of 1/2...</td>
</tr>
<tr>
<td>Collecting the Data:</td>
<td>ANPP Quadrat MeasurementsA pair of t...</td>
<td></td>
</tr>
<tr>
<td>Computer Data Entry:</td>
<td>Data is entered into an excel spreadsheet.</td>
<td></td>
</tr>
<tr>
<td>QA/QC</td>
<td>Before data are concatenated into one fil...</td>
<td></td>
</tr>
</tbody>
</table>

Study extent description. Describe the temporal, spatial and taxonomic extent of the study. This information supplements the coverage information you may have provided in a previous step.

The Sevilleta National Wildlife Refuge, operated by the U.S. Fish & Wildlife Service, is located in Socorro County in central New Mexico.

The Sevilleta spans approximately 100,000 ha of land and has vegetative influences representative of the Chihuahuan Desert, the Great Plains grassland and the Colorado Plateau.

Sampling description. Describe the sampling design of the study. For example, you might describe the way in which treatments were assigned to sampling units.

higher elevation than the Savanna study site. This site is a woodland where pinyon-pine (Pinus edulis) and juniper (Juniperus monosperma) are dominant species associated with scrub oak (Quercus turbinella), black grama (Bouteloua eriopoda), blue grama (Bouteloua gracilis), and hairy grama (Bouteloua hirsuta).
Adding contact metadata

- You can add new owner details or pick from a locally stored data package.

- Information in red is required.
### Final Product

#### Data Package: rns.150.7

**Accession Number:** rns.150.7  **Keywords:** Vertebrates, James Reserve, Species List

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual:</strong></td>
<td>Michael Hamilton</td>
</tr>
<tr>
<td><strong>Address:</strong></td>
<td>PO Box 1775, 54445 North Circle Drive,</td>
</tr>
<tr>
<td></td>
<td>Idyllwild, California 92349 USA</td>
</tr>
<tr>
<td><strong>Phone:</strong></td>
<td>9096593311</td>
</tr>
<tr>
<td><strong>Email Address:</strong></td>
<td><a href="mailto:director@jamesreserve.edu">director@jamesreserve.edu</a></td>
</tr>
<tr>
<td><strong>Individual:</strong></td>
<td>Sheri Lubin</td>
</tr>
<tr>
<td><strong>Individual:</strong></td>
<td>Kevin Browne</td>
</tr>
<tr>
<td><strong>Organization:</strong></td>
<td>James San Jacinto Mountain Reserve</td>
</tr>
<tr>
<td><strong>Organization:</strong></td>
<td>University of California Natural Reserve System</td>
</tr>
</tbody>
</table>

**Associated Party**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual:</strong></td>
<td>Michael Hamilton</td>
</tr>
<tr>
<td><strong>Address:</strong></td>
<td>PO Box 1775, 54445 North Circle Drive,</td>
</tr>
<tr>
<td></td>
<td>Idyllwild, California 92349 USA</td>
</tr>
<tr>
<td><strong>Phone:</strong></td>
<td>9096593311</td>
</tr>
<tr>
<td><strong>Email Address:</strong></td>
<td><a href="mailto:director@jamesreserve.edu">director@jamesreserve.edu</a></td>
</tr>
<tr>
<td><strong>Individual:</strong></td>
<td>Michael Hamilton</td>
</tr>
<tr>
<td><strong>Address:</strong></td>
<td>PO Box 1775, 54445 North Circle Drive,</td>
</tr>
<tr>
<td></td>
<td>Idyllwild, California 92349 USA</td>
</tr>
<tr>
<td><strong>Phone:</strong></td>
<td>9096593311</td>
</tr>
<tr>
<td><strong>Email Address:</strong></td>
<td><a href="mailto:director@jamesreserve.edu">director@jamesreserve.edu</a></td>
</tr>
<tr>
<td><strong>Individual:</strong></td>
<td>Sheri Lubin</td>
</tr>
<tr>
<td><strong>Individual:</strong></td>
<td>Kevin Browne</td>
</tr>
</tbody>
</table>
Topics

• Importance of Metadata
• Elements of Metadata
• Results of Metadata
• Morpho – Metadata Software
• **Metacat – Metadata Cataloging Software**
• KBS Task Force on Databases
• Flexible storage system for metadata and data
  • Stores metadata documents
  • Supports structured searches
  • Customizable web interface
  • Replication capabilities
The Search for Metadata

http://knb.ecoinformatics.org/index.jsp
Topics

• Importance of Metadata
• Elements of Metadata
• Results of Metadata
• Morpho – Metadata Software
• Metacat – Metadata Cataloging Software

• KBS Task Force on Databases
The KBS Task Force on Databases
The KBS Task Force on Databases

- Charged with addressing a suite of issues relating to databases within KBS.
- Inventory
- Access
- Documentation
- Dissemination
KBS Data Catalog

- Dataset name
- Description
- Keywords
- Spatial extent
- Start and end dates
- Sharing permissions
- Data source
- File name
- File Format
- Contact info
The Changing Landscape of Scholarly Communication:
The Role of Digital Repositories

- Tuesday, March 8, 2005
  9 AM - noon, Big 12 Room, Kansas Union
  1:30 - 3:00 PM, the Hall Center for the Humanities

- Sharing and preserving the products of research, including GIS data. Talks will center on digital repositories and dissemination of scholarly work, and introduce KUs own digital repository-
  ScholarWorks: www.ku.edu/~scholar/.

- Details on seminar:
  www.lib.ku.edu/scholcommSeminar.shtml.