Toward Standardization: A Participatory Framework for the Process of Developing Scientific Metadata Standards

Lynn Yarmey
National Snow and Ice Data Center
University of Colorado at Boulder

Karen S. Baker
Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois at Urbana-Champaign

LTER: Long-term Ecological Research Network

View: Site Coordinates

- Andrews Forest LTER (ANG)
- Arctic LTER (ARC)
- Baltimore Ecosystem Study (BES)
- Bonanza Creek LTER (BNZ)
- California Current Ecosystem LTER (CCE)
- Cedar Creek LTER (CCD)
- Central Arizona - Phoenix LTER (CAP)
- Coweeta LTER (CWT)
- Florida Coastal Everglades LTER (FCE)
- Georgia Coastal Ecosystems LTER (GCE)
- Harvard Forest LTER (HFR)
- Hubbard Brook LTER (HBB)
- Jornada Basin LTER (JRN)
- Kellogg Biological Station LTER (KBS)
- Konza Prairie LTER (KNZ)
- LTER Network Office (LNO)
- Luquillo LTER (LUQ)
- McMurdo Dry Valleys LTER (MCV)
- Monsoon Coral Reef LTER (MCR)
- Niwot Ridge LTER (NWT)
- North Temperate Lakes LTER (NTL)
- Palmer Antarctic LTER (PAL)
- Plum Island Ecosystems LTER (PIE)
- Santa Barbara Coastal LTER (SBC)
- Sevilleta LTER (SVL)
- Shortgrass Steppes (SGS)
- Virginia Coast Reserve LTER (VCR)
Science Studies
STS, CSCW, Historical & Infrastructure Studies

Quantitative Research  \( \rightleftharpoons \) Qualitative Research
Environmental Sciences  \( \rightleftharpoons \) Data Management Processes

Take-Home Message:
Metadata Standard-Making in the Sciences

1. Appreciating Data Differences (Inclusivity) in the standardization process

2. Establishing a Participatory Framework (Design) since standardization frameworks differ

3. Mirroring the Scientific Process (Collaboration)
Outline

1. Introduction

2. Background
   Sphere-of-Context
   Web of Repositories

3. Standard Models
   Hierarchical - Market
   Participatory - Scientific

4. Discussion
   Design
   LTER Case

5. Conclusion
   Appreciating Data Differences
   Establishing a Participatory Framework
   Mirroring the Scientific Process

In Transition...

Community Reuse

Information Management

Local Use

Data Management

Data

Local Use

Data

Millerand and Baker, 2010
Local and Remote Spheres-of-Context
Environmental Research Data Repository Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Local</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver</td>
<td>Research</td>
<td>Service</td>
</tr>
<tr>
<td>Expertise</td>
<td>Data management</td>
<td>Collection management</td>
</tr>
<tr>
<td>Design focus</td>
<td>Acquisition, capture and use</td>
<td>Storage and reuse</td>
</tr>
<tr>
<td>Data state</td>
<td>Dynamic</td>
<td>Versioned</td>
</tr>
<tr>
<td>Design feature</td>
<td>Adaptability</td>
<td>Stability</td>
</tr>
<tr>
<td>Change mechanism</td>
<td>New data types and scientific practices</td>
<td>Widely-accepted data practices</td>
</tr>
<tr>
<td>Data knowledge type</td>
<td>Tacit, implicit, and explicit</td>
<td>Explicit</td>
</tr>
<tr>
<td>Standards contribution</td>
<td>Developing and enacting</td>
<td>Propagation</td>
</tr>
</tbody>
</table>

Web of Repositories at Multiple ‘Levels’
- Federated
- Non-hierarchical
- Diverse
- Inclusive
- Distributed
- Coordinated
- Collaborative
- Flexible
- Sustainable

Baker and Yarmey, 2009
A measurement represents an observation within a specified context.

**Data Ecosystem**

The U.S. National Academy of Science (NAS, 2009) defined metadata standards as descriptions of “the content, context, and structure of information objects, including research data, at any level of aggregation (for example, a single data item, many items, or an entire database).”
Metadata Standard-Making Occurs In Many Arenas

Humanities, text oriented and specimen oriented
- TEI The Text Encoding Initiative for humanities mark up (1987)
- DC: Dublin Core Metadata Initiative for cataloguing web-based resources (1995)
- DwC Darwin Core for documenting specimen data emerged (1999)
- DDI The Data Documentation Initiative (DDI) describing social, behavioural, and economic science metadata (1997)

Natural sciences, field oriented
- FGDC Federal Geographic Data Committee (1993)
- CSDGM Content Standard for Digital Geospatial Metadata
- EML Ecological Metadata Language (1997)

LTER Case: Metadata Standard Development
Ecological Metadata Language (EML)
Design Activities

Hierarchical Framework for Developing Standards
Participatory Framework for Developing Standards

1. Individual sites
2. Negotiating
3. Interpreting

1. inclusive representation
2. Reporting
3. Negotiating

Standard

Documenting
Best Practice and Working Standard

**Best Practice**: Description of use of a standard based upon decisions made in interpreting a standard to ensure that circumstances are managed uniformly.

**Working Standard**: an ad hoc convention developed pragmatically to describe previously undescribed procedures in response to a local need.
Participatory Framework for Developing Standards

Hierarchical and Participatory Framework Characteristics

<table>
<thead>
<tr>
<th>Framework</th>
<th>Hierarchical</th>
<th>Participatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culture</td>
<td>Market</td>
<td>Research</td>
</tr>
<tr>
<td>Purpose</td>
<td>Specification</td>
<td>Integration mechanism</td>
</tr>
<tr>
<td>Goal</td>
<td>End product</td>
<td>Supporting ongoing scientific inquiry</td>
</tr>
<tr>
<td>Strategy</td>
<td>Competitive</td>
<td>Collective</td>
</tr>
<tr>
<td>Orientation</td>
<td>Solution</td>
<td>Staged cycles</td>
</tr>
<tr>
<td>Pace</td>
<td>Rapid</td>
<td>Slow</td>
</tr>
<tr>
<td>Influences</td>
<td>Politics, economics, technology</td>
<td>Existing knowledge, situated practices, technology</td>
</tr>
<tr>
<td>Dominant Driver</td>
<td>Economies-of-scale</td>
<td>Complexities-of-scale</td>
</tr>
<tr>
<td>Implementation Tools</td>
<td>Often existing</td>
<td>Often emerging</td>
</tr>
<tr>
<td>Participants</td>
<td>Limited</td>
<td>Inclusive</td>
</tr>
<tr>
<td>Standardization</td>
<td>Choosing between competing options</td>
<td>Coming to a shared understanding through learning</td>
</tr>
</tbody>
</table>
Data Ecosystem

Through process-aware data work arises the possibility of learning, designing, and sustaining the making and re-making of standards over time

Kinds of Standard-Making Organizations

National Institute of Standards and Technology (NIST) is a non-regulatory agency of the United States Department of Commerce ...

The European Information and Communications Technologies Standards Board (ICTSB) aims to coordinate standardization activities ...

Kinds: Physical, Business, IT, Geographical, Biological, etc
Kinds of Standards in the Biological Sciences

<table>
<thead>
<tr>
<th>Standard</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AALU</td>
<td><a href="http://www.biology.org/">http://www.biology.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>BRS</td>
<td><a href="http://www.bioresource.org/">http://www.bioresource.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>CONC</td>
<td><a href="http://www.consortium.org/">http://www.consortium.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>DBS</td>
<td><a href="http://www.database.org/">http://www.database.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>EMBL</td>
<td><a href="http://www.embl.org/">http://www.embl.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>GDR</td>
<td><a href="http://www.gdr.org/">http://www.gdr.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>KEGG</td>
<td><a href="http://www.genome.ad.jp/">http://www.genome.ad.jp/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>MGSA</td>
<td><a href="http://www.mgsa.org/">http://www.mgsa.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>NDB</td>
<td><a href="http://www.ndb.org/">http://www.ndb.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>PDB</td>
<td><a href="http://wwwpdb.org/">http://wwwpdb.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>PDB-CCP</td>
<td><a href="http://wwwpdb-ccp.org/">http://wwwpdb-ccp.org/</a></td>
<td>Organizational</td>
</tr>
<tr>
<td>RCSB PDB</td>
<td><a href="http://www.rcsb.org/">http://www.rcsb.org/</a></td>
<td>Organizational</td>
</tr>
</tbody>
</table>

Kinds of Standards Organizations

1. Coordinating Unit
   - Community
   - Consortium
   - Government
   - Market

2. Standards Object
   - Process
   - Semantics
   - Performance
   - Product

Thessen and Patterson (2011)
In Conclusion: Standard-Making

- **Appreciating Data Differences (Inclusivity):** Collaborative work needed at many levels and many spheres-of-context to avoid metadata shoehorned into ill-fitting standards to meet minimum requirements.

- **Establishing a Participatory Framework (Design):** Standard-making in the sciences is at best an ongoing, participatory process.

- **Mirroring the Scientific Process (Collaboration):** In its provisionally, the process of standardization parallels the ongoing nature of the scientific knowledge-making process itself ... continuing and collaborative.

---

Thank You

Essentials for the Process of Standardization in the Sciences

- Appreciating Data Differences
- Establishing a Participatory Framework
- Mirroring the Scientific Process