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

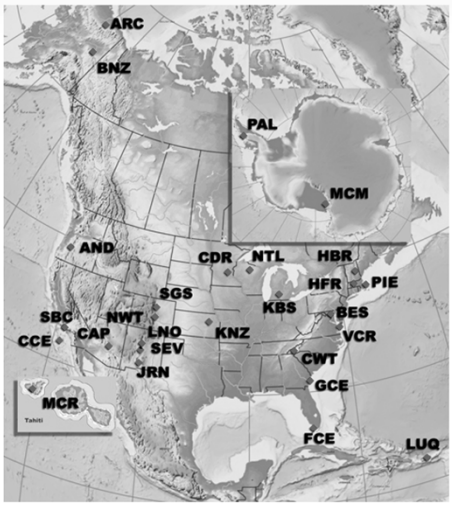
Lynn Yarmey
National Snow and Ice Data Center
University of Colorado at Boulder

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

1

LTER: Long Term Ecological Research Network



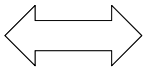
View: Site Coordinates

- Andrews Forest LTER (AND)
- Arctic LTER (ARC)
- Baltimore Ecosystem Study (BES)
- Bonanza Creek LTER (BNZ)
- California Current Ecosystem LTER (CCE)
- Cedar Creek LTER (CDR)
- 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 (HBR)
- Jornada Basin LTER (JRN)
- Kellogg Biological Station LTER (KBS)
- Konza Prairie LTER (KNZ)
- LTER Network Office (LNO)
- Luquillo LTER (LUQ)
- McMurdo Dry Valleys LTER (MCM)
- Moorea Coral Reef LTER (MCR)
- Niwot Ridge LTER (NWT)
- North Temperate Lakes LTER (NTL)
- Palmer Antarctica LTER (PAL)
- Plum Island Ecosystems LTER (PIE)
- Santa Barbara Coastal LTER (SBC)
- Sevilleta LTER (SEV)
- Shortgrass Steppe (SGS)
- Virginia Coast Reserve LTER (VCR)

2

Science Studies
STS, CSCW, Historical & Infrastructure Studies

Quantitative Research
-
Environmental Sciences



Qualitative Research
-
Data Management Processes

3

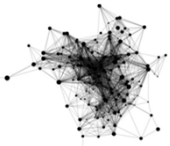
Take-Home Message:
Metadata Standard-Making in the Sciences

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)

4

Outline

From the field

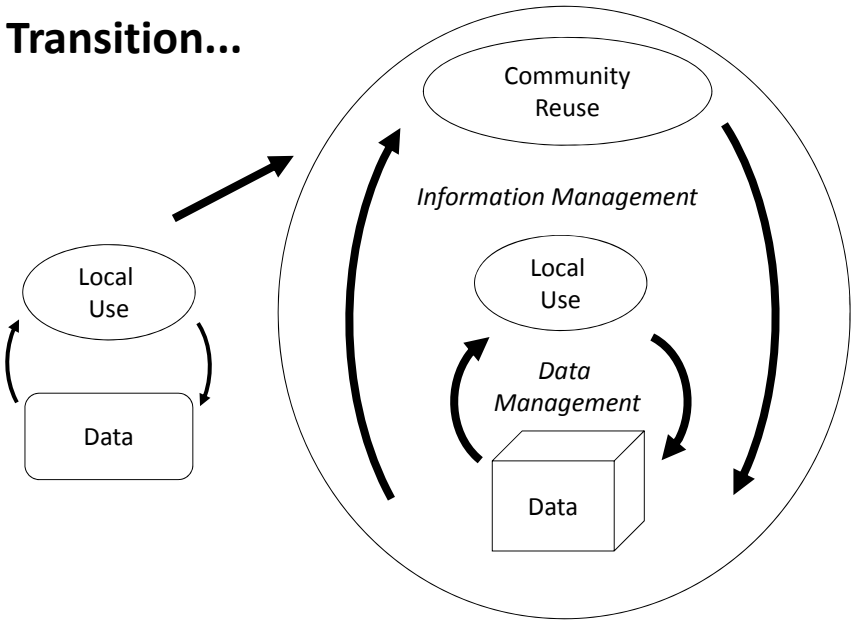


to the web
of data repositories

- ✓ 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

5

In Transition...



6

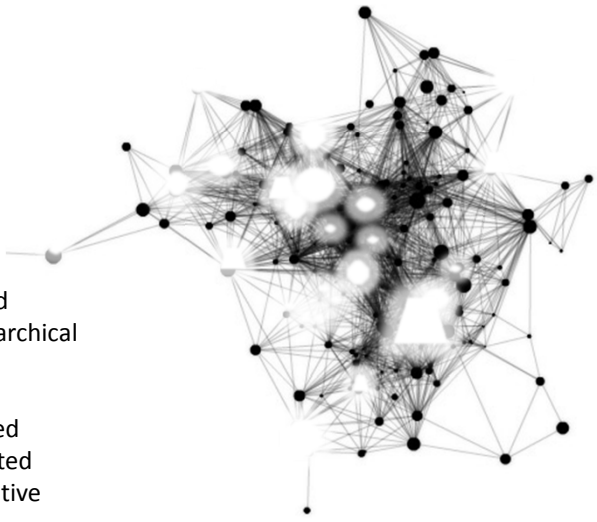
Millerand and Baker, 2010

Local and Remote Spheres-of-Context Environmental Research Data Repository Characteristics

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

7

Web of Repositories at Multiple 'Levels'

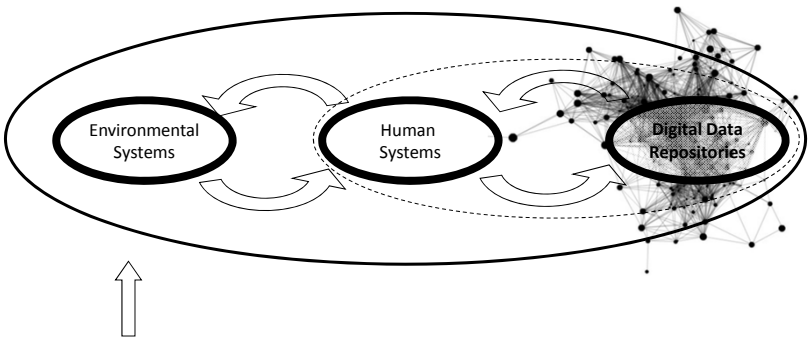


- Federated
- Non-hierarchical
- Diverse
- Inclusive
- Distributed
- Coordinated
- Collaborative
- Flexible
- Sustainable

Baker and Yarmey, 2009

8

Data Ecosystem

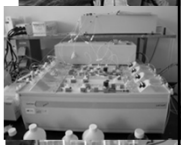


A measurement represents an observation within a specified context.

9

Baker and Chandler, DSR, 2008

Standard-Making Metadata in the Sciences



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

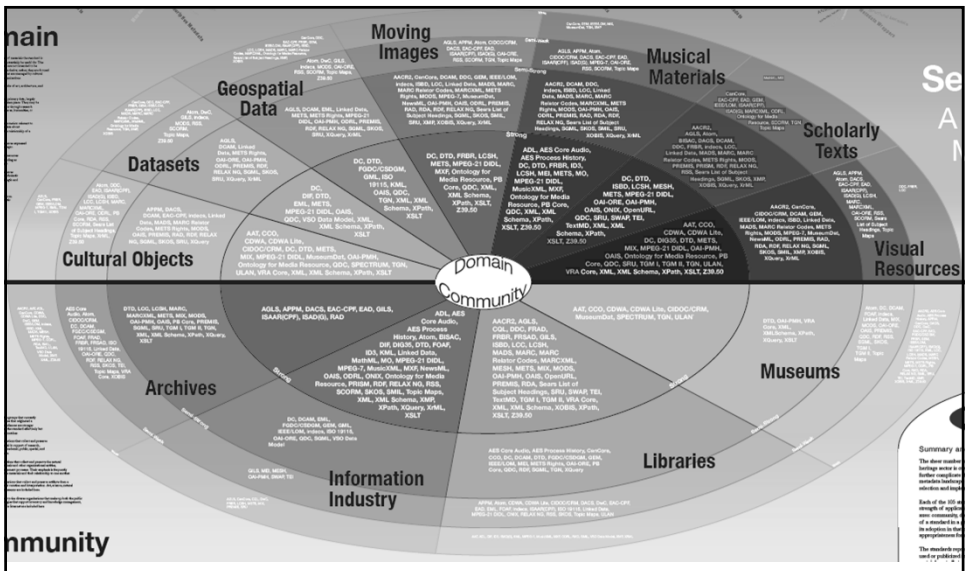
Humanities, text oriented and specimen oriented

- 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

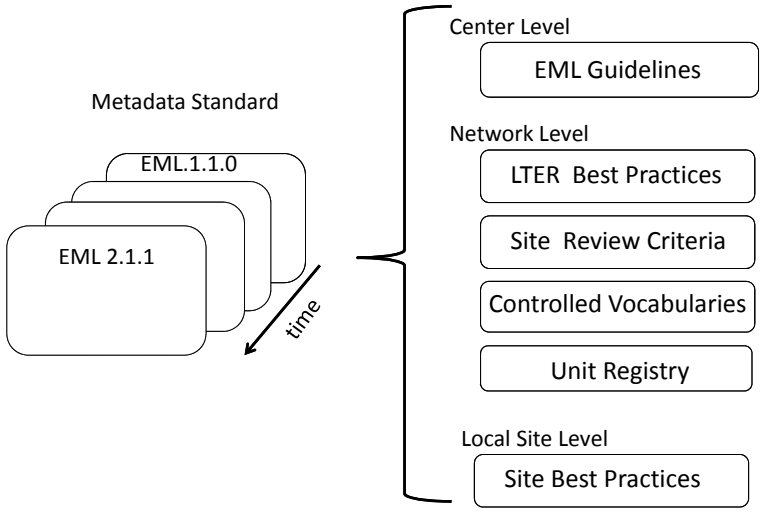
- FGDC Federal Geographic Data Committee (1993)
  CSDGM Content Standard for Digital Geospatial Metadata
- EML Ecological Metadata Language (1997)

11

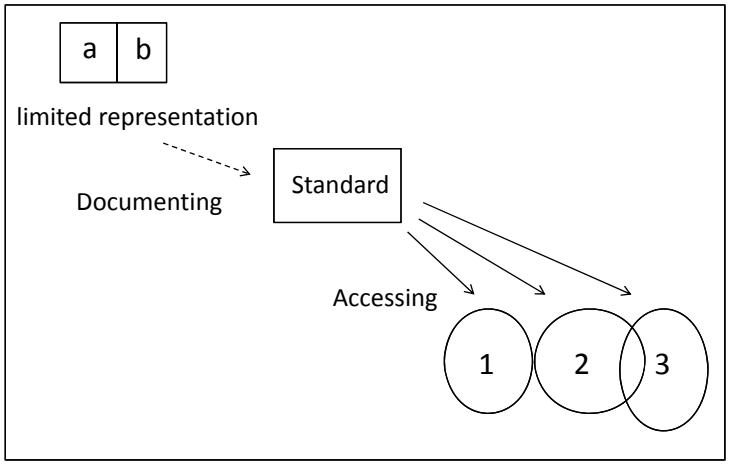


Riley, J., 2010. Seeing Standards: A Visualization of the Metadata Universe.http://www.dlib.indiana.edu/~jenlrile/metadatamap/

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

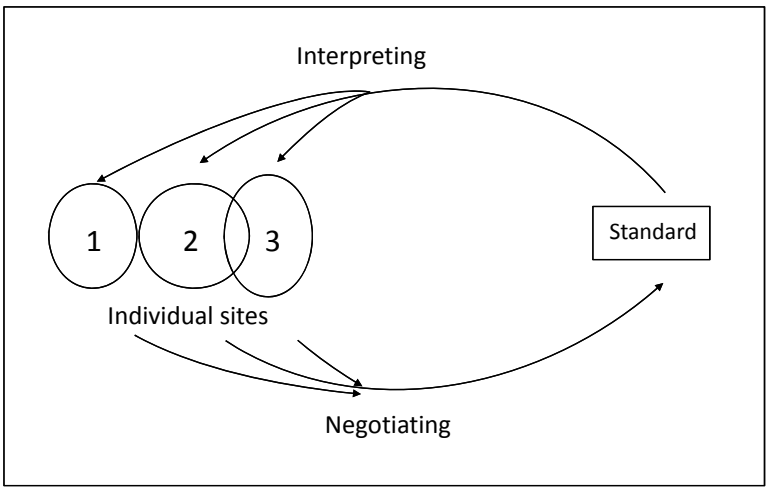


Hierarchical Framework for Developing Standards



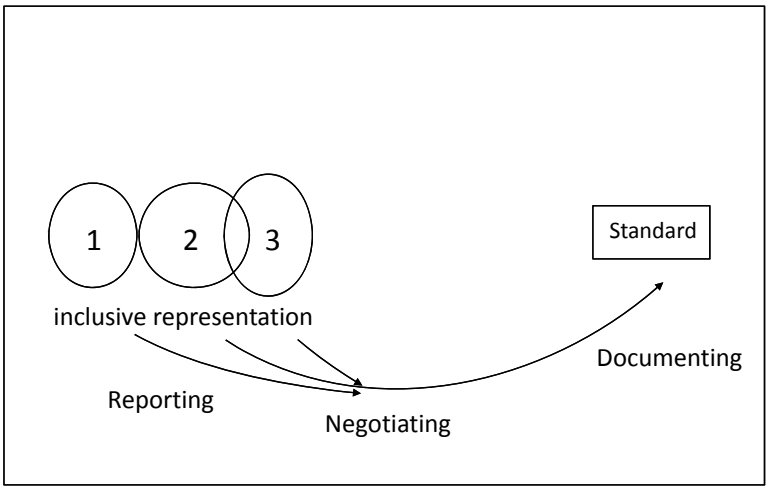
14

Participatory Framework for Developing Standards



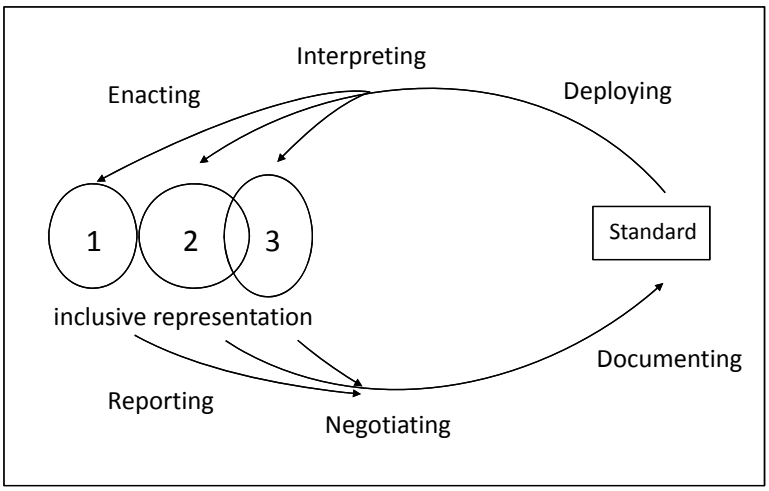
15

Participatory Framework for Developing Standards



16

Participatory Framework for Developing Standards



17

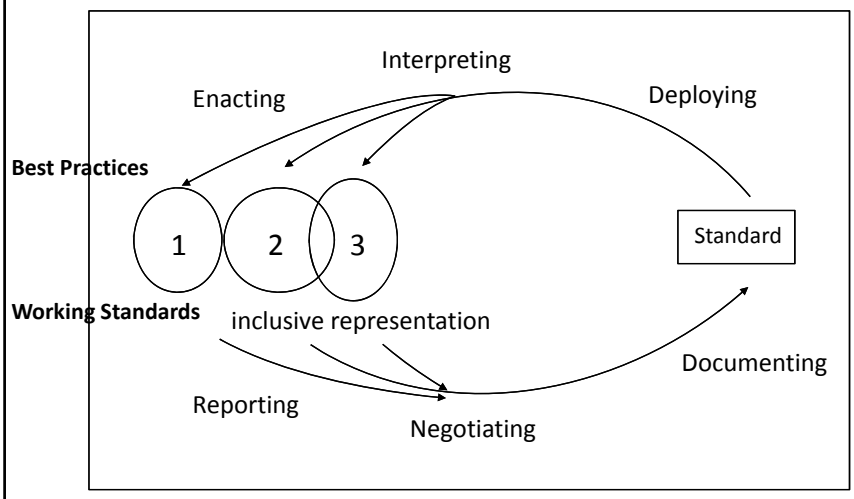
Best Practice and Working Standard

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.

18

Participatory Framework for Developing Standards



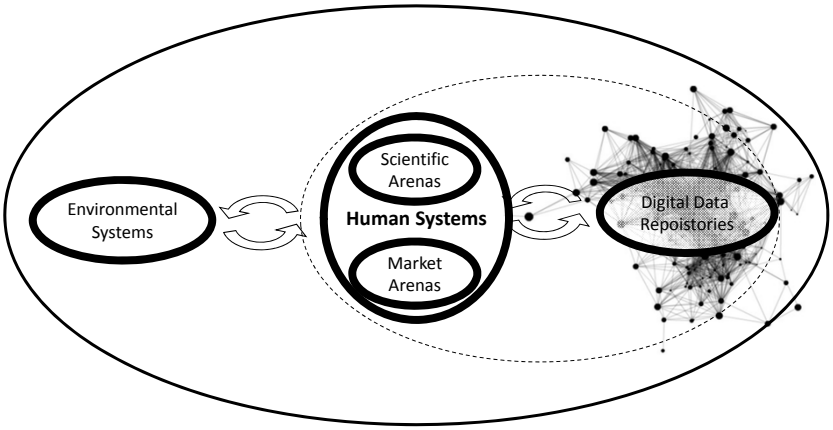
19

Hierarchical and Participatory Framework Characteristics

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

20

Data Ecosystem



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

21

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

22

Kinds of Standards in the Biological Sciences

Standard	Location	Type
ABCD	http://www.bgbm.org/TDWG/CODATA/Schema/default.htm	Schema
Bioontology	http://www.bioontology.org/	Ontology Repository
BIRN	http://www.birncommunity.org/	
Cardiac Electrophysiology Ontology	http://bioportal.bioontology.org/ontologies/39038	Ontology
CMECS	Coastal and marine ecological classification standard http://www.csc.noaa.gov/benthic/cmecs/cmecs_doc.pdf	Vocabulary
Comparative Data Analysis ontology	http://sourceforge.net/apps/mediawiki/cdao/index.php?title=Main_Page	Ontology
Darwin Core	http://wiki.tdwg.org/twiki/bin/view/DarwinCore/	Metadata
Dublin Core	http://dublincore.org/	Metadata
Ecological Metdata Language	http://knb.ecoinformatics.org/software/eml/	Metadata
Environment Ontology	http://www.environmentontology.org/	Ontology
Evolution Ontology	http://code.google.com/p/evolution-ontology/	Ontology
Experimental Factor Ontology	http://www.ebi.ac.uk/efo/	Ontology
Federal Geospatial Data Committee	http://www.fgdc.gov/	Metadata
Fungal Anatomy	http://www.yeastgenome.org/fungi/fungal_anatomy_ontology/	Ontology
Gene Ontology	http://www.geneontology.org/	Ontology
Homology Ontology	http://bioportal.bioontology.org/ontologies/42117	Ontology
HUPO	http://www.psidev.info/index.php?q=node/159	Vocabulary
Infectious Disease ontology	http://www.infectiousdiseaseontology.org/Home.html	Ontology
International Standards Organization	http://www.iso.org	Metadata
Marine Metadata Interoperability	http://marinemetadata.org/	Metadata
Miriam	http://www.ebi.ac.uk/miriam/main/datatypes/	Vocabulary
National Biodiversity Information Infrastructure	http://www.nbii.gov/portal/community/Communities/NBII_Home/	Metadata
Ontology of Microbial Phenotypes	http://sourceforge.net/projects/microphenotypes/	Ontology
Open Biological and Biomedical Ontologies	http://www.obofoundry.org/	Ontology Repository
Phenotype Quality Ontology	http://obofoundry.org/wiki/index.php/PATO:Main_Page	Ontology
Plant Ontology	http://www.plantontology.org/	Ontology

Thessen and Patterson (2011)

23

Kinds of Standards Organizations

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

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

U. Lower (2010)

24

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.

25

Thank You

Essentials for the Process of Standardization in the Sciences

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

26