2440

Research Metadata in Long-Term Ecological Research (LTER)

John H. Porter
Virginia Coast Reserve, LTER
Department of Environmental Sciences
University of Virginia
Clark Hall
Charlottesville, VA USA 22903
phone: (804) 924-8999
fax: (804) 982-2137
jporter@lternet.edu

Donald L. Henshaw
H.J. Andrews Experimental Forest LTER
USDA Forest Service
Pacific NW Station
3200 SW Jefferson Way
Corvallis, OR USA 97331
phone: (541) 750-7335
fax: (541) 750-7329
dhenshaw@lternet.edu

Susan G. Stafford
H.J. Andrews Experimental Forest LTER
Oregon State University
Department of Forest Science
Forestry Sciences Lab. 020
Corvallis, OR USA 97331-7501
phone: (541) 750-7335
fax: (541) 750-7329
sstafford@lternet.edu

ABSTRACT

Long-Term Ecological Research (LTER) Sites manage a diverse array of ecological information resulting from research on ecological succession, disturbance, landscape ecology, elemental cycling, trophic structure, biodiversity, organic matter and primary productivity. Since its inception in 1980, the LTER Network has evolved from a primarily site-based research focus to a network and cross-site focus. Coincident with this evolution has been the development of metadata standards. Site-specific metadata standards and systems are being integrated through the development of metadata exchange standards and the revolution in network information servers.

1.0 INTRODUCTION

Eighteen research sites and a Network Office make up the U.S. Long-term Ecological Research (LTER) Network. Focusing on fundemental ecological research, each LTER site is supported by an individual grant from the National Science Foundation. The LTER sites are found in forest, desert, prairie, alpine, coastal and marine environments. The research topics addressed at sites are similarly varied, with research on ecological succession, disturbance, landscape ecology, elemental cycling, trophic structure, biodiversity, organic matter and primary productivity ongoing at each

site. The sites also differ in their degree of centralization, with researchers from up to 15 universities at a single site (Hayden, 1996).

The types of data collected and used at LTER sites are highly variable. Satellite images, scanned aerial photos, and output from spatially explicit models fall at the high-end of size, with an individual image requiring over 300 MB of disk space. At the other end of the spectrum are types of data, such as deep soil cores where a small number of samples are subjected to a large number of different tests. This may involve metadata that are be orders of magnitude larger than the associated data files.

The diversity of the LTER sites, the scientific questions being addressed, the data being collected and the administrative structures under which they operate pose a significant challenge for information management efforts (Stafford et. al., 1994). The approach taken by the LTER Network since its inception in 1980 has been for each site to operate its own information management system under a data management policy that conforms to LTER Network guidelines, with the LTER Network Office taking the lead in linking those individual systems. This allows each site to focus on the types of data most prevalent at that site by choosing software and hardware compatible with the mission and the computational and administrative environments of the site.

2.0 EVOLUTION OF LTER METADATA

At its inception in 1980, the sites receiving funding from the NSF LTER program focused on scientific questions that were primarily site-specific. Site information and metadata management systems were designed to meet the needs of scientists within a single sites. The focus was on providing the metadata required to extract data stored in electronic forms, primarily ASCII files. Metadata could be relatively specific, because scientists working at each site were generally knowlegible about the site conditions and objectives. Nonetheless, the metadata had to be sufficient to assure that the data would be interpretable over the long term. The goal was to produce metadata that would allow the data could be used 20-years subsequent to its collection.

However, starting with the addition of sites in 1987, there has been an increasing focus on cross-site and network activities, although site-based research remains at the core of the LTER program. This broadening of the program placed a premium on broadening of the metadata as well. The users of the data could no longer be assumed to be knowlegible about the basics of a site. Additionally, exchange of data between sites became increasingly important. Starting in 1994, LTER sites received a mandate to place data online from the LTER Coordinating Committee (the governing group of U.S. LTER). This broadened metadata needs yet again.

Table 1: Milestones in the evolution of LTER Information Management Systems

Year	Milestone
1980	NSF funds first cohort of LTER sites
1986	"Research Data Management in the Ecological Sciences" book (Michener, 1986) based on research experiences at LTER sites
1988	LTER Network Office established
1990	LTER Guidelines for Data Access formulated (Porter and Callahan, 1994)
1990	LTER Core Dataset Catalog published
1992	Sites establish network-accessible information servers
1993	LTER All-Site Bibliography assembled (Chinn and Bledsoe, 1997)
1994	Mandate for LTERs to make data available online approved by LTER Coordinating Committee
1994	LTER Information Managers commit to developing Network Information System
1994	LTER Content Standards for Metadata Exchange formulated
1996	Metadata catalog prototype development

3.0 STANDARDS DEVELOPMENT

With the growth of the LTER Network (as compared to the initiation of individual LTER sites), development of

standards has become increasingly important. Despite their importance, the approach to standards has been iterative and cautious, with a focus on emergent, rather than imposed, standards. Scientific research, by its nature, requires a constant state of change. The identification of new questions and creation of new ways of measuring nature are an inherent part of the scientific enterprise. Premature imposition of standards before the state-of-the-art has sufficiently advanced can constrain a research community, just as lack of standards can frustrate attempts at synthetic and integrative research. Formalization of emergent standards provides a way of balancing those disparate needs.

Sites in the LTER Network act as experimental units, each evaluating a variety of methods at the site level. We therefore took a comparative approach in defining guidelines for metadata exchange. Sites were asked to send copies of site metadata standards to a committee of site information managers. Those managers then integrated the lists, finding those elements that were common across sites. These constituted a viable minimum metadata content standard. In addition, there were some elements that were not widely used, but that had clear utility. They were included as optional elements in a draft metadata content standard (Kirchner et al., 1994).

The draft standard was further developed by adding elements that were needed for cross-site cataloging efforts, but were not necessary for site-specific metadata (e.g., within a site, it is not necessary to include the name of the site). The resulting draft standard was then compared with other evolving standards, notably the Federal Geographic Data Committee Standards for Geospatial Metadata, the Global Change Master Directory's Directory Interchange Format (DIF) (NASA, 1983) and the Biological Resources Division of the US Geological Survey (then the National Biological Service). The LTER draft content standard has been subsequently used as a component in other metadata development efforts (Michener et al., 1997).

Table 2: LTER Metadata Content Standards. Uppercase terms are equivalent to FGDC and NBS (now USGS/BRD) terms (adapted from Kirchner et al., 1994).

Dataset information

- DATASET IDENTITY - name or title by which the dataset is known (type text, domain free text)
- Type of data [fixed field, delimited field, binary, image, etc.)
- ORIGINATOR - name of the organization(s) or individual(s) that developed the dataset. If the names of editors or compilers are provided, they must be followed by (ed.) or (comp.)
    - Originator address
    - Originator voice telephone
    - Originator facsimile telephone
    - Originator electronic mail address
- TIME PERIOD OF CONTENT (start and stop dates)
- Dissemination restrictions
- AVAILABLE TIME PERIOD or RELEASE DATE (NBS)
- DATASET DESCRIPTION - a description of the dataset including its intended use and limitations (type text, domain free text)
- CONTACT INFORMATION
    - CONTACT PERSON
    - CONTACT MAIL ADDRESS
    - CONTACT VOICE TELEPHONE
    - CONTACT FACSIMILE TELEPHONE
    - CONTACT ELECTRONIC MAIL ADDRESS
- KEYWORDS -words or phrases summarizing some aspect of the dataset (type compound)
    - THEME KEYWORD (required)
    - PLACE
    - STRATUM
    - TEMPORAL
- IDENTIFICATION CODE1 -unique item or stock code by which the item could be ordered or the full path name to the file (type text, domain free text, N/A, UNKNOWN)
- DATASET CITATION or CITATION_INFORMATION1 (preferred)
- SPATIAL DOMAIN1 (bounding coordinates, or arbitrary polygon)

- Sample storage1 (Information on storage of samples)
- RELATED DOCUMENTS1- Reference citations and location information
- Attribute (variable) documentation
    - ATTRIBUTE LABEL Type [integer, floating point, character/string]
    - Format
        - Start column
        - End column
        - Optional number of decimal places
    - ATTRIBUTE DEFINITION
    - ATTRIBUTE UNITS OF MEASUREMENT
    - ATTRIBUTE DOMAIN VALUES - CODE SET DOMAIN
    - ENUMERATED DOMAIN VALUE DEFINITION (list of coded values)
    - RANGE DOMAIN
        - RANGE_DOMAIN_MINIMUM
        - RANGE_DOMAIN_MAXIMUM
    - Missing value codes1 (should be required if such codes can appear in the data)
    - Precision2
- Methods of collection

Project Information

- Project description
- Project Title
- Date commenced
- Date terminated or Expected duration
- Objectives
- Abstract
- Source of funding
- Principal investigator
- Additional investigators
- Responsible investigators/technicians/supervisor

Site description

- Site type (terrestrial, aquatic stream, aquatic lake, etc.)
- Watersheds
- Permanent plots
- Habitat
- Soil
- Slope/aspect
- Terrain/physiography
- Geology/lithology
- Hydrology
- Size
- History
- Elevation
- Climate

Dataset description

- Dataset title
- Number of records
- Data form used
- Location of completed data forms
- Literature
- Method of recording
- Associated computer accounts

- Comments
- LTER core area
- Dataset files (subsets)
- Citation
- Treatment of data (programs used in analysis, plus reference/link to their metadata)
- Date of last review
- Date of last entry
- Researcher review status
- Taxa/functional group
- Statistical analyses Supporting datasets
- Other supporting materials
- Maps
- Aerial photos
- Digital images
- GPS
- Use history
- Request histories
- Update histories

4.0 CONCLUSIONS

The "bottom line" for information management within LTER is the diversity - both in terms of the types of ecological information and the specificities and eccentricities of the individual sites. This diversity provides opportunities for experimentation in information management, but also intrinsically imposes certain contraints on our ability to manage and share information within and among the individual sites. Given the rapid creation rate of new ecological databases and an increasing number of cooperative, intersite information comparitive efforts, it is imperative that we find better ways to facilitate data production and data integration of these diverse databases (Spycher et al., 1996). Consequently, recent LTER efforts have focused on the development of database catalogs and exchange standards that will serve as building blocks in the development of the generic tools necessary for data integration.

5.0 REFERENCES

[1] B.P. Hayden "An LTER Profile: Fifteen years old, the LTER model is showing its worth" LTER Network News, Issue 19 p. 1, Spring/Summer 1996. Also available at: http://lternet.edu/about/program/table.htm

[2] S.G. Stafford, J.W. Brunt and W.K. Michener "Integration of scientific information management and environmental research" In Environmental Information Management and Analysis: Ecosystem to global scales, eds. S.G. Stafford, J.W. Brunt and W.K. Michener Taylor & Francis, Bristol, PA pp. 3-19, 1994.

[3] Michener, W.K (ed.). 1986. Research Data Management in the Ecological Sciences. Belle W. Baruch Library in Marine Science No. 16, University of South Carolina Press, Columbia, SC. p. 426.

[4] Porter, J.H. and J.T. Callahan. 1994. Circumventing a dilemma: historical approaches to data sharing in ecological research. In W.K. Michener, S. Stafford and J.W. Brunt (eds.). "Environmental Information Mangagement," Taylor and Francis, Bristol, PA pp. 193-203.

[5] Chinn, H. and C. Bledsoe. 1997. Internet access to ecological information -- the US LTER All-site Bibliography Project. Bioscience 47:50-57.

[6] Kirchner, T., H. Chinn, D. Henshaw and J. Porter. 1994. Documentation Standards for Data Exchange. In: Proceedings of the 1994 LTER Data Management Workshop. R. Ingersoll and J. Brunt (eds.). gopher://lternet.washington.edu:70/11/newsletters/Reports/DMworkshops/1994.

[7] Directory Interchange Format Manual, Version 4.1, World Data Center A For Rockets and Satellites 92-20, NASA Goddard Space Flight Center, April 1993. http://gcmd.gsfc.nasa.gov

[8] Michener, W.K., J.W. Brunt, J.J. Helly, T.B. Kirchner and S.G. Stafford. 1997. Non-geospatial metadata for the ecological sciences. Ecological Applications 7(1):330-342.

[9] Spycher, G.; Cushing, J. B.; Henshaw, D. L.; Stafford, S. G.; Nadkarni, N. 1996. Solving problems for validation, federation, and migration of ecological databases. In: Global networks for environmental information: Proceedings of Eco-Informa '96. 1996 November 4-7; Buena Vista, FL. 11: 695-700. Ann Arbor, MI, Environmental Research Institute of Michigan (ERIM).

ACKNOWLEDGEMENTS

This paper would not have been possible without the efforts of scientists and information managers at each of the LTER sites. It was supported by NSF Grant DEB-9411974 to the University of Virginia and NSF Grant DEB-9632921 to Oregon State University.

Paper Index
Second IEEE Metadata Conference

September 16-17, 1997
NOAA Auditorium, NOAA Complex
Silver Spring Metro Center
1301 East-West Highway
Silver Spring, Maryland

Sponsored by:

IEEE Computer Society Mass Storage Systems Technology Technical Committee
National Oceanic and Atmospheric Administration
Hughes Information Technology Systems
National Imagery and Mapping Agency

In cooperation with IEEE Computer Society Digital Libraries Task Force

---

Search

1. New Approaches to Cataloguing, Querying and Browsing Geospatial Metadata

   Leon Shklar , Edith Au
   Pencom Web Works, 40 Fulton St., New-York, NY, USA 10038

   Cliff Behrens , Chumki Basu
   Bellcore, 445 South St., Morristown, NJ, USA 07960

---

2. Interoperability Issues of the European Catalogue of Data Sources (CDS)

   Ralf Kramer , Ralf Nikolai , Claudia Rolker
   Forschungszentrum Informatik (FZI), Haid-und-Neu-Strasse 10 - 14, D-76131 Karlsruhe, Germany

   Sigfus Bjarnason
   European Environment Agency (EEA), Kongens Nytorv 6, DK-1050 Copenhagen K, Denmark,

   Stefan Jensen
   Niedersechsisc hes Umweltministerium, Referat 606 (Lower Saxony Ministry of Environment), Archivstrasse 2, D - 30169 Hannover, Germany

---

3. Metadata to Support Information Based Computing Environments

   Chaitanya Baru , Richard Frost , Richard Marciano , Reagan Moore , Arcot Rajasekar , Michael Wan
   Enabling Technologies Group, San Diego Supercomputer Center, 10100 Hopkins Dr., La Jolla, CA, USA 92093

---

4. Research Metadata in Long-Term Ecological Research (LTER)

   John Porter
   Virginia Coast Reserve, LTER, Department of Environmental Sciences, University of Virginia, Clark Hall,

Charlottesville, VA USA 22903

Donald Henshaw
H.J. Andrews Experimental Forest LTER, USDA Forest Service, Pacific NW Station, 3200 SW Jefferson Way, Corvallis, OR USA 97331

Susan Stafford
H.J. Andrews Experimental Forest LTER, Oregon State University, Department of Forest Science, Forestry Sciences Lab. 020, Corvallis, OR USA 97331-7501

---

5. NASA Earth Observing System Data and Information System Implementation of Extra- and Inter- File Metadata: Application to Data Services

L. Klein , S.J.S. Khalsa, J. Gales, D. Wynne
Space Applications Corp., 901 Follin Lane, Suite 400, Vienna, Virginia 22180

B. Fortner
Hughes STX Corp., 7701 Greenbelt Rd., Greenbelt, Maryland 20770

D. Ilg
Fortner Research LL.C., 100 Carpenter Rd., Sterling, Virginia 20164

---

6. Multimedia Metadata - our 70 year experience

Chris Wilkie
British Broadcasting Corporation, Information and Archives Department, Reynard Mills, Windmill Rd, Brentford, Middx UK TW8 9NQ

---

7. Building A Statistical Metadata Repository

Daniel Gillman , Martin Appel , Samuel Highsmith
Bureau of the Census, Statistical Research Division, SRD Rm.3000-4, Washington, DC, USA 20233-9100

---

8. Institutions, Languages, Content Standards, Record Syntaxes, and Protocols: Fitting It All Together

Robert Bourdeau , Cheryl Burley , Karen Collamore Sullivan , Gary Roseberry
Consortium for International Earth Science Information Network, 2250 Pierce Rd., University Center, MI, USA 48710

---

9. Distributed Active Relationships in the Warwick Framework

Ron Daniel Jr.
Advanced Computing Lab, Los Alamos National Laboratory

Carl Lagoze
Computer Science Department, Cornell University,

---

10. Meta-Information Models for Georeferenced Digital Library Collections

Kate Beard
Department of Spatial Information Science and Engineering, University of Maine, Orono, ME 04469