Metadata for Long-Term Ecological Research

Kristin Vanderbilt
kristin.vanderbilt@orst.edu
Department of Forest Science
Oregon State University
Corvallis, OR 97331, USA

Long-term ecological data are a valuable resource for scientists addressing questions at broad temporal and spatial scales. The interdisciplinary nature of research about global change, biodiversity, and ecosystem sustainability often requires that scientists utilize data collected by other researchers. Physical, biological, and chemical data sets that share a common spatial or temporal domain can be synthesized to model complex ecological systems. In addition, as new research questions arise long-term data are being reused for purposes other than those for which they were originally collected. Because human memory is short, it is vital that long-term data sets be thoroughly documented as they are collected to prevent information loss and to ensure that secondary users of the data can reanalyze it correctly.

To facilitate data reuse and sharing, data sets must be accompanied by detailed information, or **metadata**, that describe the context, content, quality, structure, and availability of the data set. Without metadata, a data set consists only of columns of raw data that are of little value. Metadata are the information that describe 1) what data are in a database, 2) how the data were collected, 3) the objectives of the researcher collecting the data, 4) the scale relevance of the data, 5) how the data set is structured, and 6) other relevant information that may affect secondary use of the data.

Long-term ecological data present special challenges for data documentation. Long-term experiments often go through changes in sampling procedures, instrumentation, personnel, disturbances to the study system such as flooding or fire, or other fluctuations in environmental conditions. Records of these and other factors that cause changes in the data are critical if the ecological significance of the data is to be correctly interpreted. A systematic and standardized way to document and preserve such details is needed.

While metadata standards exist for geospatial data (Federal Geographic Data Committee 1994), no such standard presently exists for non-geospatial data. Michener et al. (1997), however, have proposed a set of generic metadata descriptors that could serve as a standard for long-term ecological data. This system includes five categories of descriptors (Table 1). Category I descriptors include basic attributes of the data set that indicate the temporal and spatial scales of the data, and an abstract describing the study objectives. Category II descriptors include information about the study design, methods of data collection, and the personnel involved in the study. Category III metadata indicate if the data set is available for secondary use, and how recently the data set and metadata have been updated. Class IV metadata describe the structure of the data file, including variable names, variable definitions, and missing value codes. Class V metadata document information related to the data set that may be helpful to a scientist reusing the data, such as publications based on the data set, problems with the data detected by other users, or the availability of voucher specimens.

Documentation of all five categories of metadata descriptors for every data set is clearly impossible due to time and financial constraints. The extent of metadata development depends on who the anticipated secondary data users may be (Michener et al. 1997). If data exchange will occur with expert colleagues, then only Level I and Level IV descriptors are needed. If metadata are to be useful to a broader audience, then Level I through IV metadata are recommended. Level V metadata is always desirable.

Table 1. Proposed metadata descriptors and examples (based on Michener et al. 1997)

Descriptors	Examples
Class I. Data Set Descriptors	
a. Data set identity	Title of data set
b. Data set identification code	Unique identifying code for data set
c. Data set description	Investigator names and addresses; Abstract summarizing research objectives
d. Key words	Location and temporal scale of data
Class II. Research Origin Descriptors	
a. "overall" project description	Identity, originator(s), period of study, objectives of study, descriptive abstract, source of funding
b. "specific subproject" description	Site description (geography, habitat, geology, hydrology, climate, site history), experimental design, research methods, project personnel
Class III. Data set status and accessibility	
a. Status	Latest update, latest archive date, metadata status
b. Accessibility	Storage location, contact persons, proprietary restrictions
Class IV. Data structural descriptors	
a. Data set file	File name, size, format
b. Variable information	Variable names, definitions, units of measurement, range, precision, missing value codes
c. Data anomalies	Description of missing data, calibration errors
Class V. Supplemental Descriptors	
a. Data acquisition	Examples of data forms, digitizing procedures
b. Quality assurance/Quality control	Treatment of outliers, equipment performance
c. Related materials	References and location of maps, photographs, GIS layers
d. Publications	
e. History of data set usage	Log of who used data and for what purpose, comments from secondary users

References

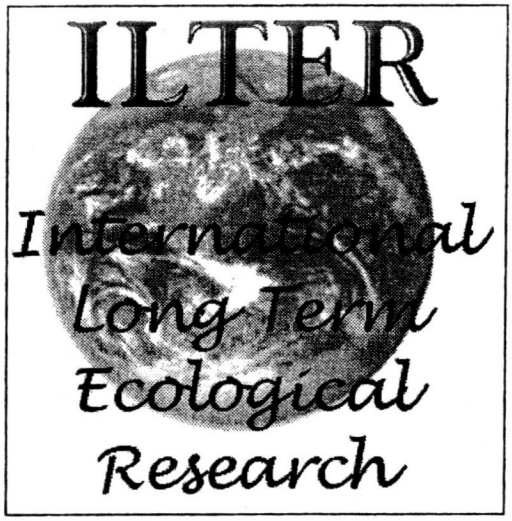
Federal Geographic Data Committee. 1994. Content standards for digital geospatial metadata (June 8). Federal Geographic Data Committee, Washington, D.C., USA.

Michener, William K., Brunt, James W., Helly, John J., Kirchner, Thomas B. and Stafford, Susan G. 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications 7:330-342.

Cooperation in Long Term Ecological Research in Central and Eastern Europe

Proceedings of the ILTER Regional Workshop
22-25 June, 1999
Budapest, Hungary

Edited by Kate Lajtha and Kristin Vanderbilt
Oregon State University



Cover design: Janet Squire

Cover illustration: The polluted regions of the rivers studied in the Eastern-tributaries of the Tisa River. Figure 1 from "The ecological state of the Eastern-tributaries of the Tisa River - based on characteristics of the physico-chemical parameters, the flora and fauna" by Andrei Sárkány-Kiss and Kunigunda Macalik.

Printed by Oregon State University, Printing and Mailing Services, January 2000

Lajtha, K. and K. Vanderbilt, eds. 2000. Cooperation in Long Term Ecological Research in Central and Eastern Europe: Proceedings of the ILTER Regional Workshop, 22-25 June, 1999, Budapest, Hungary. Oregon State University, Corvallis, OR.