

Ecological and Environmental Data

Gerry McKiernan

With the widespread development and adoption of the Internet and World Wide Web over the past several years by specialists as well as the public, an increasing number of government agencies, publishers, and researchers have adopted these technologies to provide access to primary and secondary ecological and environmental data or to sources that describe or interpret them. While many of these sources are well-known, a number are less conventional or have been recently introduced, and thus, are not presently widely used. In this *Sci-5* review, we profile lesser-known data sources of potential interest to ecologists, environmentalists, and others.

CITE: Ecological Society of America. Committee on the Future of Long-Term Ecological Data. *Final Report of the Ecological Society of America Committee on the Future of Long-Term Data (FLED)*. 2 vols. Washington, D.C.: Ecological Society of America, 1995. Also available at <<http://esa.sdsc.edu/FLED/FLED.html>> (1 November 2001).

Gerry McKiernan, AB, MS, is Associate Professor, Science and Technology Librarian, and Bibliographer, Iowa State University Library, Ames, IA (E-mail: gerrymck@iastate.edu).

[Haworth indexing entry note]: "Ecological and Environmental Data." McKiernan, Gerry. Published in *Science & Technology Libraries* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 23, No. 4, 2003, pp. 95-104. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

<http://www.haworthpress.com/web/STL>

© 2003 by The Haworth Press, Inc. All rights reserved.

Digital Object Identifier: 10.1300/J122v23n04_08

* * *

WHAT? CEED: 'Caveat Emptor' Ecological Data Repository

WHERE? <http://ceed.sdsc.edu/>

WHEN? 1999

WHY? CEED: 'Caveat Emptor' Ecological Data Repository contains data associated with peer-reviewed journal articles as well data contributed from non-peer-reviewed data collection projects, and is intended to facilitate collaborative ecological scientific research.

HOW? Although registration is not required to search and browse file information and metadata within CEED, pre-registration is recommended before initiating a search. The Web-based registration form requests basic contact information (i.e., first, middle and last name, organizational affiliation, street mailing address, daytime phone number, e-mail address) and a user-supplied password. After processing the submission, the system assigns a user identification number ('User ID') (e.g., u00001). After formal registration, users can search CEED and subsequently review metadata for a data file and indirectly its contents. The CEED catalog can be searched by author (first and/or last name), title or title keywords, or subject keywords, from within a general search screen. If desired, a search may be limited to an associated time period ('Begin Date' and 'End Date').

Users can also search CEED for relevant data files for a given geographic region. This search is performed by pointing, holding, and dragging a mouse cursor to outline a desired geographic area within an interactive world map. The field search and interactive geographic search can be used concurrently. Presently, data files in CEED relate primarily to southern California organisms and environments.

Upon executing the search ('Query'), the file identifier, 'Metadata Title,' and file 'Size' in kilobytes are provided for each relevant file separate listing (e.g., 'u00010.19990218.1737.19990220.1811.tar.gz,' 'San Diego Bay Birds,' and '82515,' respectively).

Within the current CEED system, users do not directly access selected data files from within the CEED catalog but indirectly through an FTP (File Transfer Protocol) address associated with the individual CEED record. To retrieve the full file and its contents, the user must first click the associated check box located to the left of each file number in the listing. After finishing the check-off of relevant files, a 'Done' button, found at the bottom of the screen, is clicked. The user is then presented with a form requesting his or her user identification and the associated password; upon completion, the user clicks another 'Done' but-

ton. Upon successful login, the screen presents a standard message indicating that the requested search is being processed, notifying the user that he/she will receive an e-mail message when the requested files are ready to be downloaded; the e-mail will typically be sent within minutes of the initial request. After receiving notification, a file may be downloaded through a browser. Once downloaded, it must be uncompressed and the archived data file extracted using an appropriate program (e.g., WinZip).

To view the metadata associated with a file, the user clicks the *metadata title*, in the original file listing. The standardized CEED record includes numerous types of fields and associated data. Among the field types and sub-types are Dataset (Dataset, Subject, Coverage, Rights, Status) Files (File, Variables), Research (Project, Site), and Supplemental (Supplemental). Field and sub-field record data are accessible by expanding nodes on a 'Metadata Tree' displayed on the left side of the screen. In total, there are more than 50 data fields in all.

Users with relevant data are invited to contribute to the CEED repository and explicit instructions are provided to assist in uploading files. General guidance is also offered for completion of the requisite metadata fields and sub-fields.

WHO? Principal Investigator: John J. Helly (hellyj@sdsc.edu), Senior Research Scientist, San Diego Supercomputer Center, University of California, San Diego. The project is supported by a grant from the National Science Foundation (BIR/DBA/DBI-9631091) at the San Diego Supercomputer Center.

CITE: John J. Helly, T. Todd Elvins, Don Sutton, and David Martinez, "A Method for Interoperable Digital Libraries and Data Repositories," *Future Generation Computer Systems* 16, no. 1 (November 1999): 21-28.

J. J. Helly, T. T. Elvins, D. Sutton, and D. Martinez, S. Miller, S. Pickett, A. M. Ellison, "Controlled Publication of Digital Scientific Data," *Communications of the ACM*. Forthcoming. Available at: <http://www.sdsc.edu/~hellyj/papers/jjh02_acm.pdf> (2 November 2001).

* * *

WHAT? *Ecological Archives*

WHERE? <http://www.esapubs.org/archive/>

WHEN? 1997

WHY? *Ecological Archives* (EA) is a public-available, free Web site that publishes materials supplemental to articles that appear in the print journals of the Ecological Society of America (ESA) (*Bulletin, Ecology, Ecological Ap-*

plications, and *Ecological Monographs*) as well as the peer-reviewed *Data Papers*. The ability to publish appendices and supplements in EA allows authors to prepare a more concise version of a submission for publication in paper. EA provides access to a wide variety of supporting material that includes data tables, graphs illustrating additional analyses, photographs, additional references, as well as supplemental discussion and methodological details.

HOW? *Data Papers* (DP) are compilations and syntheses of datasets and their associated metadata considered to be of potential interest to members of the ESA and the scholarly community. They differ from other review or synthesis papers published in ESA journals in that DPs typically do not test or refine ecological theory. DPs are subject to full peer-review and oversight by subject-matter editors and a Data Editor. The review process is intended to evaluate the ecological significance and overall quality of a submission as well as the usability of its associated metadata and its adherence to the *EA Metadata Content Standard* [<http://ecodata.sdsc.edu/metadata/example.html>]. The full titles, author(s), abstracts, and keywords of peer-reviewed DPs are published in *Ecology* or *Ecological Applications* as appropriate, along with *Ecological Archives* accession numbers.

Prior to the established of EA, ESA maintained a *Supplemental Publication Service* (ESPS) from 1981 through 1997 to provide permanent access to article supplements published in ESA journals. The 68 sets of supplemental information contained in the ESPS files have been incorporated within EA and are treated as supplements.

The ESA provides long-term maintenance of a publicly accessible archival copy of a DP and accepts periodic updates. To offset the cost associated with providing continued access, authors are charged a one-time fee for DP publication. Currently, the DP fee is \$250 for publication of the abstract in the relevant journal and storage of the DP in EA to 10MB. To encourage authors to submit succinct contributions to its journals and monograph series, ESA does not charge for the publication in EA for appendices or supplemental materials that cumulatively total less than 10 MB. For DPs, appendices, or supplemental files greater than 10MB, additional fees are charged.

Permission to make digital or hard copies of part or all of the contents of *Ecological Archives* for personal or educational use within an individual's home institution is granted without fee, provided that the first page or initial screen of a display includes the notice "Copyright by the Ecological Society of America." In addition, the full citation, including the name of the author(s) of the EA contribution must be included.

WHO? *Ecological Archives* was established by Robert K. Peet (peet@unc.edu), The University of North Carolina at Chapel Hill, during his tenure as editor-in-chief of *Ecology* and *Ecological Monographs* [<http://www.bio.unc.edu/faculty/peet/service.htm>]. The current EA Editor is William K. Michener (wmichene@lternet.edu), Senior Research Scientist, United States Long-Term Ecological Research Network, Department of Biology, University of New Mexico, Albuquerque, NM and the current EA Manager is Jane L. Bain (jlb40@cornell.edu), Data Archives Manager, ESA, Cornell University.

CITE: Ecological Archives. Index [http://www.esapubs.org/archive/archive_index.htm].

* * *

WHAT? *Ecological Information Network*

WHERE? <http://ein.nbii.gov/>

WHEN? 2000/2001

WHY? The *Ecological Information Network* (EIN) is an online, publicly available electronic directory that provides biographical and professional information about ecological experts who have volunteered to respond to queries on various scientific issues. The EIN is open to all scientists with expertise in any area relevant to the field of ecology.

HOW? The EIN directory can be searched using a 'Simple Search' or an 'Advanced Search' interface. The former allows the user to enter a single key to search the directory; the latter provides drop-down menus that direct the user to 'Expertise Categories' and 'Expertise Subjects.'

The 'Advanced Search' is a two-step process in which the user first selects from one of several 'Expertise Categories' in a drop-down menu and then selects from a list of 'Expertise Subjects' associated with a selected category from an adjacent drop-down menu. The following are the 'Expertise Categories':

- Ecological Subdiscipline Expertise
- Ecological Expertise
- EcoSystem Expertise
- Ecological Zone of Expertise
- Expertise in Ecological Education/Diversity Issues

For a specific category, there are associated subject expertise subcategories. For example, the subject areas associated with the 'Ecological Zone of Expertise' has the following associated general subject subcategories:

- Arctic Tundra
- Boreal Forest and Taiga
- Deserts and Xeric Shrublands
- Flooded Grasslands and Savannas
- Mangroves
- Marine Ecoregion
- Mediterranean Shrublands and Wetlands

There are two identical sets of drop-down menus for both the 'Expertise Categories' and the 'Expertise Subjects' allowing the user to combine two selected expertise subjects in an 'OR,' 'AND,' or 'NOT' Boolean relationship.

In executing a search, a list of registered experts is presented in alphabetical order by the individual's surname; adjacent to each name is the individual's institutional affiliation. For example:

The following folks have working knowledge (2) or expertise (3) in "Boreal Forests and Taiga"

(2) Foley, Patrick X California State University Sacramento

(2) Kellogg, Laurie E University of Notre Dame

(2) Vadas, Jr., Robert (Bob) L. Washington Department of Fish and Wildlife.

The full name is hotlinked to a brief database record for the individual and provides basic identification (first, middle, last names) and essential contact information (e.g., full address and e-mail). This record, in turn, is hotlinked to an extensive and comprehensive record that provides additional data about the individual, detailed information about his or her area(s) of expertise, and his or her willingness to assist members of the profession, the public, or teachers, among other activities.

To assist users in searching using the basic or advanced search options, access is provided to an *Expertise Subject Quick Reference Page* <<http://ein.nbii.gov/quikref.htm>> from within the Advanced Search page (*Quick Reference List of Expertise Categories and Subjects*).

WHO? The Ecological Information Network is a joint project of the Ecological Society of America (ESA) and the National Biological Information Infrastructure (NBII). The EIN site was developed and maintained by the Center for Biological Informatics of the U.S. Geological Survey.

Individuals interested in registering with EIN can do so by completing an online questionnaire ('Sign up to be an Expert') [<http://ein.nbii.gov/quest.htm/>].

CITE: Lori Hiding and Ron Slangen, "Connecting ESA and NBII: The Ecological Information Network." Presentation at *Connecting Ecological Information: National Biological Information Infrastructure and Ecologists* workshop held during the annual meeting of the Ecological Society of America, Snowbird, Utah, August 7, 2000 <<http://www.nbii.gov/about/presentation/esa/ein.html>> (29 October 2001).

* * *

WHAT? *Environmental Data Registry*

WHERE? <http://www.epa.gov/edr/>

WHEN? EDR 6.0 (September 14, 2001)

WHY? The *Environmental Data Registry* (EDR) is a catalog of the major data collections of the United States Environmental Protection Agency (EPA). As such, it serves as a comprehensive, authoritative source of reference information about the definition, source, and uses of environmental data.

HOW? As a reference tool, the EDR catalog provides information for interpreting EPA environmental data. The EDR does not contain the environmental data itself, but rather information that describes it (metadata). As the repository for standard data elements, the EDR facilitates the use of consistently defined and formatted data elements and sets of data values, fostering information consistency through standard information representation. The EDR contains EPA data standards and business rules documents, specifications for standard and candidate standard data elements, standard code sets for use as permissible values in application systems, EPA application system metadata including table structure and data element specifications and values, and news about the EPA data standard program. As of October 2001, EDR contained nearly 9,000 (8,823) data elements and held nearly 1,600 (1,594) information sources, contributed by 52 submitting organizations.

From within the ERS homepage, users may:

- Access information about approved and candidate data standards designed to promote efficient sharing of environmental information;
- Use the Search functions to find and download metadata elements by source, by keyword, or by organization;
- Select data elements or standard data values (code sets) to download for use in application system design and reengineering; and

- Access related components of an integrated metadata registry, a reference to Environmental Protection Agency (EPA) information resources. Related systems include the *Substance Registry System* (SRS), the *Chemical Registry System* (CRS), the *Biology Registry System* (BioRS), the *Terminology Reference System* (TRS), and *MetaPro Metadata Registry Builder*.

In addition, access is provided to the *Standard Update Newsletter*, a quarterly that covers topics related to data standards (e.g., “Metadata—Understanding Its Meaning and Uses”; “The New Integrated EDR”; and “Standards Stewardship and Implementation”). Issues of the newsletter are available in HTML, Word 97, WordPerfect 8, and PDF formats. EDR users may also subscribe to a free news and e-mail alerting service.

EDR is based on the International Organization for Standardization (ISO) *Information Technology. Specification and Standardization of Data Elements* (ISO/IEC 11179:1999) <<http://www.diffuse.org/meta.html#ISO11179>> and the American National Standards Institute (ANSI) for Information Technology *Meta-model for the Management of Shareable Data* (ANSI X3.285-1998) <<http://www.sdct.itl.nist.gov/~ftp/x318/other/Standards/ANSIX3.285-1998.pdf>>, an elaboration of ISO 11179:1999.

WHO? EDR was developed by the Environmental Protection Agency as a repository for standard data elements to simplify the exchange of environmental data between the agency and affiliated organizations. EDR is available to anyone interested in environmental information, including the public, environmental decision-makers, industry, researchers, regulated entities, state and local environmental agencies, regulation writers, and computer system developers and managers.

CITE: Environmental Protection Agency (U.S.). *Environmental Data Registry: Frequently Asked Questions*. Available at: <[http://oaspub.epa.gov/edr/faq\\$.startup](http://oaspub.epa.gov/edr/faq$.startup)> (27 October 2001).

* * *

WHAT? *National Geospatial Data Clearinghouse*

WHERE? <http://fgdclearhs.er.usgs.gov/>

WHEN? 1996?

WHY? The National Geospatial Data Clearinghouse is a collection of over 250 spatial data servers that offer digital geographic data primarily for use in geographic information systems (GIS), image processing systems, and other

modeling software. The primary intention of the clearinghouse is to provide discovery services for digital data allowing users to evaluate their quality through metadata descriptions.

HOW? Users may access and search the clearinghouse servers after selecting one of six clearinghouse gateways from an image map. Although the search options are not identical, each gateway provides access to the same collection of servers. From within the gateway provided by the Federal Geographic Data Committee (FGDC), the clearinghouse can be searched using a *Search Wizard* to 'smart select' servers and data, a Java applet-based map interface with place names, or a place name that does not use a Java applet.

Using the *Search Wizard*, users can search by choosing from twenty-one (21) broad topics (e.g., 'Biologic and Ecologic Information'). With this interface, the user may choose all topics, or a maximum of four topics. After selection, the user can limit a search by place name or geographic area. State names or country names maybe selected from a defined scrollable list, while geographic areas can be selected by specifying the geographic coordinates of the area or by outlining the area on an interactive map. In the subsequent step, the user receives an alphabetical listing of all of the servers that meet the conditions of the query. Standardized information about each server is available from an associated hotlink found adjacent to the server name. Servers from this listing can be reviewed and deselected if considered inappropriate for a given search.

From within the next search screen, users can further limit a search to a specific time period ('Specify Time Period of Content') and/or specify a specific field (i.e., 'Any,' 'Title,' 'Originator,' 'Abstract,' 'Purpose,' 'Presentation Form' or 'Theme Keyword') in which to execute a canned search strategy. A predefined search strategy (e.g., 'biolog*' or 'bioinf*' or 'ecolog*' or 'habitat or ecosys*' or 'biome') is used to represent the content of the selected topic(s) (e.g., 'Biologic and Ecologic Information'), but can be modified by the user and/or refined by selecting specific search fields and/or Boolean combinations with additional terms or phrases.

After initiating the search, each candidate server is contacted and the status of the connection shown as the search is executed through the server sub-collection. When completed, the results from individual servers can be assessed by clicking on the highlighted hotlinked server name, if results were found within a specific server. For matching files within a selected server, the title of the file, as well as a link to a brief record ('Summary') containing the name, time period, and geographic coverage and a full metadata record ('Full') for the digital resource. Except for interactive mapping Web services made available from within specific servers, most digital resources accessible from within the clearinghouse are not ready-to-use maps, but the raw spatial data bases that can generate them.

WHO? National Geospatial Data Clearinghouse is a component of the National Spatial Data Infrastructure (NSDI) and is sponsored by the Federal Geographic Data Committee, an U.S. interagency committee, organized to coordinate the use, sharing, and dissemination of geospatial data on a national basis.

CITE: Douglas Nebert, "Building a Geospatial Data Clearinghouse for Data Discovery and Access," Paper presented at *Work Session on Methodological Issues Involving the Integration of Statistics and Geography*, Conference of European Statisticians, Neuchâtel, Switzerland, April 11, 2000 <<http://www.unece.org/stats/documents/2000/04/gis/8.e.pdf>> (2 November 2001).

* * *

NOMINATIONS

Members of the science and technology community are invited to nominate quality science and technology Web sites and resources for potential review in *Sci-5*. Of greatest interest are sites with uncommon but useful content, and those with innovative features and functionalities. Nominations should be sent to Gerry McKiernan (gerrymck@iastate.edu).