



---

# CONTRIBUTIONS

---

## A Brief History of Data Sharing in the U.S. Long Term Ecological Research Network

In the abstract, the advantages of sharing data are manifest. No individual scientist, or even small group of scientists, can collect all the data that are needed to address today's major ecological research questions, especially those dealing with global, regional, or long-term phenomena. Recently, the need for scientists of all disciplines to share data has been highlighted. The 10 September 2009 issue of *Nature* featured a cover story, and editorials focusing on sharing data (Anonymous 2009, Authors 2009, Nelson 2009, Schofield et al. 2009, Toronto International Data Release Workshop Authors 2009). The integration of data from diverse sources opens new opportunities for research to the common benefit of the ecological research community and humankind in general (National Research Council 1997, Porter 2000, Anonymous 2006, Michener 2006, Wicherts et al. 2006). However, as in many scientific disciplines, data sharing remains more the exception than the rule throughout most of ecology (Wolins 1962, Craig and Reese 1973, Costello and Berghe 2006, Costello 2009, Nelson 2009).

Among the exceptions is the U.S. Long Term Ecological Research (LTER) Network, which now provides access to >6000 individual data sets. However, this was not always so. In its early days, the LTER data was as tightly held as any ecological data today. Tracking the changes in policies and attitudes toward data sharing within the LTER Network can help provide insights into how ecologists as a whole can benefit from sharing data.

The LTER Network is not the only group that has been working to promote sharing of ecological data. (Arzberger et al. 2004) lay out a set of operating principles for data access. The National Center for Ecological Analysis and Synthesis has played critical roles in developing tools to support data sharing (Jones et al. 2001, Fegraus et al. 2005, Jones et al. 2007, Borer et al. 2009). The Ecological Society of America committee on the Future of Long-term Ecological Data (FLED) helped cement standards for ecological metadata (Michener et al. 1997). The Organization of Biological Field Stations, professional societies, and various governmental entities such as the NASA Distributed Active Archive Centers (Cook et al. 2001) and the NOAA National Ocean Data System are important sources of ecological data. These organizations (and others) have all made substantial contributions. However, here I focus on how a single organization, the LTER Network, went from virtually no data sharing to ubiquitous data sharing.

In the late 1980s, researchers within the LTER Network were no more likely to share data than were

---

other researchers. For example, during a 1989 visit to discuss possible connections to the Internet (then newly open to ecological researchers), the lead principal investigator of one LTER site forcefully stated (literally banging his fist on his desk for emphasis): “If being connected to the Internet means people can get access to our data, we don’t want it!” His concerns were somewhat ameliorated by a discussion of Internet security (few of us would want to expose the entire contents of our computers to the world), but he had remaining concerns about “stealing data” or being “scooped” on publications.

To address these concerns, at the 1990 LTER All-Scientists Meeting, an ad hoc committee was convened. With substantial input from the data managers and investigators at the LTER sites, the group stopped short of recommending a network-wide policy for data sharing, and instead developed guidelines for information policies at individual sites (Table 1).

In addition to the guidelines, they also included a sample policy, addressing four different classes of data, that would meet the requirements of the guidelines (Table 2).

Why did the committee adopt guidelines and not a policy? A major reason was that, at that time, the whole notion of sharing data was foreign to most of the researchers in the LTER network. A dictatorial policy could have driven researchers away from the network, fearing that their data was being “stolen.” Moreover, researchers’ active participation in the preparation of documentation (metadata) would be required if the data were to be usable by others (Michener et al. 1997). A second reason was that ecological researchers in general had little experience in crafting data-sharing policies. It was not known what would work and what wouldn’t. By adopting guidelines rather than a policy, the committee engaged the researchers at each LTER site in the design of guidelines that would work, at least at the level of the individual site.

Researchers can’t ask for data they don’t know exists, and through the 1980s it was very hard to even learn what sorts of data sets were being maintained by any LTER site. This changed in 1990, which saw the first version of an LTER-wide data catalog in printed form (Michener et al. 1990). This catalog consisted of summary descriptions of data sets (but not the data themselves) for a minimum of 10 “core” data sets from each of the LTER sites. Even this catalog was an ambitious step in a time where both sites and researchers clung closely to their data. However, once created, it helped alert the research community to the data resources and their potential to address a wide array of intersite questions.

By 1993, the majority of LTER sites had site-specific data-sharing policies in place. The resulting individual site policies were reviewed by (Porter and Callahan 1994). A simple additive model, based on the time devoted to different tasks, was used to examine “evolutionarily stable strategies” for promoting data sharing. One conclusion was that sharing data in the absence of policies that mandated attribution was unsound. Individuals who share their data need to be rewarded, either through receiving scientific credit via acknowledgement, citation or co-authorship, or by receiving financial remuneration (e.g., royalties, increased likelihood of future grant funding). Not surprisingly, many of the LTER site data policies echoed the sample policy (Table 2), with the identification of different classes of data and allowing data collectors to withhold data for designated time periods.

During the period 1991 through 1993, a revolution in information technology occurred with the

Table 1. LTER guidelines for site data management policies (1990).

Each Long Term Ecological Research (LTER) site should develop its own data management policy in consultation with key investigators and higher administrative units. The following provides general guidelines and rationale, but each site should be prepared to defend its own policy through the site and peer review process. The general policy of the Division of Biotic Systems and Resources, National Science Foundation (NSF), is that the data are public property one year after termination of the relevant grant.

#### General Guidelines

The management policy should include provisions that assure:

- ☐ The timely availability of data to the scientific community;
- ☐ That researchers and LTER sites contributing data to LTER databases receive adequate acknowledgement for the use of their data by other researchers and that sites receive copies of any publication using that data;
- That documentation and transformation of data are adequate to permit data to be used by researchers not involved in its original collection;
- That data must continue to be available even if an investigator leaves the project through transfer or death;
- That standards of quality assurance and quality control are adhered to;
- That long-term archival storage of data is maintained;
- That researchers have an obligation both to contribute data collected with LTER funding to the LTER site database and to publish the data in the open literature in a timely fashion;
- That costs of making data available should be recovered directly or by reciprocal sharing and collaborative research;
- That LTER data sets not be resold or distributed by the recipient; and
- That investigators have a reasonable opportunity to have first use of data they collected.

---

Table 2. Sample policy meeting the criteria listed in the guidelines (1990).

The following is an example of a policy that will meet these guidelines with respect to data sharing.

**Data Type I. Published data and metadata (i.e., data about data).**

Policy: Data are available upon request without review.

**Data Type II. Collective data of the LTER site (usually routine measurements generated by technical staff).**

Policy: Data are available for specific scientific purposes one year after generation.

**Data Type III. Original measurements by individual researchers.**

Policy: Data are available for specific scientific purposes two years after generation.

Data can be released earlier with permission of the researcher.

**Data Type IV. Unusual long-term data collected by individual researchers.**

Policy: The principal investigator of the LTER site can designate that such data can be withheld for longer periods. Such action should be rare and justified in writing.

advent of Internet-based online systems (FTP, Gopher, and subsequently the World Wide Web). The 1990 guidelines and the site data-sharing policies did not explicitly address the online sharing of data. In 1990, only one LTER site operated a dialup “bulletin board” type system, but by 1993 most of the sites had online capabilities in one form or another. With substantial urging by the National Science Foundation, in 1994 the LTER Coordinating Committee mandated that, by the end of the year, each site should have at least one data set available online.

The trickle of online data sets that started in 1994 turned into a flood in subsequent years, as LTER sites competed with one another in seeing who could make the most information available. Reviewers of renewal proposals increasingly looked at the value of the data being made available by an LTER site in deciding whether to continue funding. For this reason, the adoption in 1997 of the simplified network-wide data-sharing policy (Table 3) was relatively simple.

Table 3. Data access policy for the LTER Network (1997). <<http://lternet.edu/data/netpolicy.html>>

- 1) There are two types of data: Type I (data that are freely available within 2–3 years) with minimum restrictions, and Type II (Exceptional data sets that are available only with written permission from the PI/investigator(s)). Implied in this timetable, is the assumption that some data sets require more effort to get online and that no "blanket policy" is going to cover all data sets at all sites. However, each site would pursue getting all of their data online in the most expedient fashion possible.
- 2) The number of data sets that are assigned TYPE II status should be rare in occurrence and that the justification for exceptions must be well documented and approved by the lead PI and site data manager. Some examples of Type II data may include: locations of rare or endangered species, data that are covered by copyright laws (e.g. TM and/or SPOT satellite data) or some types of census data involving human subjects.

What has been the result of this policy over the last decade? Costello (2009) lists a variety of concerns that discourage researchers from sharing data, including: plagiarization, being “scooped” on papers, commercial use, and lack of recognition. The experience of the LTER network in the intervening period is that sharing data has done a great deal of good and that instances of abuse have been extremely rare. How rare? An e-mail query was sent to all the Information Managers at the LTER sites. In aggregate, those who responded reported on the results of 31,789 data set downloads and identified a grand total of four instances where problems occurred: (1) where a litigator requested unpublished data for court room use, (2) where a data requestor lied about their identity (circumstantial indications are that it was a K–12 student), (3) different researchers downloaded the same data to work on similar papers without knowing that the other was doing so, and finally (4) where a researcher disagreed with a subsequent interpretation of their data. Taken together these problems occurred in <0.1% of the requests. This low rate of problems undermines many of the concerns listed by Costello (2009). Moreover, they have led to system modifications such as the addition of disclaimers about the suitability of preliminary data for forensic use, improved systems for authenticating data users, and logging systems that inform data users about who else is interested in the data and might be worth contacting as a potential collaborator. Only in the last instance (disagreement on interpretation of data) is it not possible to design data system elements to help ameliorate the potential for disputes, and a strong case can be made that such disagreements are a fundamental part of the scientific process and we should not attempt to eliminate them.

In most data exchanges, the users of the data are more than happy to give proper credit to the providers of the data in their publications, or even to include them as co-authors. Instead of facilitating “stealing” of data (e.g., misrepresenting someone else’s data as your own), online systems with time-stamped metadata containing “author” information provide added protection for collectors of data, since online systems provide clear evidence of priority. One interesting observation is that frequently users of published data sets are the investigators who collected the data originally, or their students, because working with well-documented data from a structured archive is easier than tracking down poorly or undocumented files on your own computer.

---

The broad acceptance of data sharing by the LTER community was made manifest, when attendees at a January 2001 meeting of the lead investigators from each of the LTER sites were revising the statement of goals for the network. Only a single completely new goal was added: “Information: To inform the LTER and broader scientific community by creating well-designed and well-documented databases.” <http://www.lternet.edu/research/goals.html>

What are some of the lessons learned during this process? The first is that it is important to engage the researchers or institutions that will be providing data in the formulation of data policies. Only if they are comfortable with the provisions in the policy will they contribute. A second lesson is that all the “stakeholders” in the data-sharing process have responsibilities. The data collector is responsible for collecting high-quality data and supporting metadata, and for providing access to it. The manager of the online system has the responsibility for archiving and preserving the data and metadata and to make sure that it is only accessible to users within the context of the policy (e.g., authenticated users). Finally, and perhaps most importantly, the data user has the responsibility for properly acknowledging or citing their use of the data.

The final lesson is that it is not necessary to do it all at once. A critical aspect of building confidence in the concept of sharing data was the stepwise approach, which allowed data contributors to develop a level of comfort at each stage before moving to the next step.

## Acknowledgments

Members of the LTER Information Management Committee participated in innumerable discussions of issues regarding data sharing, and without their work, this paper would not have been possible. Donald Henshaw, John Anderson, and Jonathan Walsh provided information on data use and related problems, and William Michener and Susan Stafford provided constructive comments on the manuscript. This work was supported by NSF Grant DEB-0621014.

## Literature cited

- Anonymous. 2006. A fair share. *Nature* 444:653–654.
- Anonymous. 2009. Data’s shameful neglect. *Nature* 461:145. [doi:10.1038/461145a]
- Arzberger, P., P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhler, and P. Wouters. 2004. An international framework to promote access to data. *Science* 303:1777–1778.
- Borer, E. T., E. W. Seabloom, M. B. Jones, and M. Schildhauer. 2009. Some simple guidelines for effective data management. *ESA Bulletin* 90:205–214. <http://www.esajournals.org/doi/abs/10.1890/0012-9623-90.2.205>
- Cook, R. B., R. J. Olson, P. Kanciruk, and L. A. Hook. 2001. Best practices for preparing ecological data sets to share and archive. *ESA Bulletin* 82:138–141. <http://www.jstor.org/stable/20168543>
- Costello, M. J. 2009. Motivating online publication of data. *BioScience* 59:418–427. <http://www.bioone.org/doi/abs/10.1525/bio.2009.59.5.9>



- Costello, M. J., and E. V. Berghe. 2006. 'Ocean biodiversity informatics': a new era in marine biology research and management. *Marine Ecology Progress Series* 316:203–214. <[http://www.int-res.com/articles/meps\\_oa/m316p203.pdf](http://www.int-res.com/articles/meps_oa/m316p203.pdf)>
- Craig, J. R., and S. C. Reese. 1973. Retention of raw data: a problem revisited. *American Psychologist* 28:723.
- Fegraus, E. H., S. Andelman, M. B. Jones, and M. Schildhauer. 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. *ESA Bulletin* 86:158–168. <<http://www.esajournals.org/doi/abs/10.1890/0012-9623%282005%2986%5B158%3AMTVOED%5D2.0.CO%3B2>>
- Jones, C., C. Blanchette, M. Brooke, J. Harris, M. Jones, and M. Schildhauer. 2007. A metadata-driven framework for generating field data entry interfaces in ecology. *Ecological Informatics* 2:270–278. <<http://www.sciencedirect.com/science/article/B7W63-4PGPM4V-1/2/8915b247926b83716286bfcd7df235d9>>
- Jones, M. B., C. Berkley, J. Bojilova, and M. Schildhauer. 2001. Managing scientific metadata. *IEEE Internet Computing* 5:59–68. <[http://www.icess.ucsb.edu/lter/biblio/finalms/jones\\_mb2001pgs59-68pdf.pdf](http://www.icess.ucsb.edu/lter/biblio/finalms/jones_mb2001pgs59-68pdf.pdf)>
- Michener, W. K. 2006. Meta-information concepts for ecological data management. *Ecological Informatics* 1:3–7. [doi: 10.1016/j.ecoinf.2005.08.004]
- 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* 7:330–342.
- Michener, W. K., A. B. Miller, and R. W. Nottrott, editors. 1990. Long-Term Ecological Research Network core data set catalog. Belle W. Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, South Carolina, USA.
- National Research Council. 1997. Bits of power: issues in global access to scientific data. National Academy Press, Washington, D.C., USA.
- Nelson, B. 2009. Data sharing: empty archives. *Nature* 461:160–163. [doi:10.1038/461160a]
- Porter, J. H. 2000. Scientific databases. *in* W. K. Michener and J. Brunt, editors. *Ecological data: design, processing and management*. Blackwell Science, London, UK.
- Porter, J. H., and J. T. Callahan. 1994. Circumventing a dilemma: historical approaches to data sharing in ecological research. Pages 193–203 *in* W. K. Michener, S. Stafford, and J. W. Brunt, editors. *Environmental information management*. Taylor and Francis, Bristol, Pennsylvania, USA.
- Schofield, P. N., T. Bubela, T. Weaver, L. Portilla, S. D. Brown, J. M. Hancock, D. Einhorn, G. Tocchini-Valentini, M. H. d. Angelis, N. Rosenthal, and CASIMIR Rome Meeting participants. 2009. Post-publication sharing of data and tools. *Nature* 461:171–173.
- Toronto International Data Release Workshop Authors. 2009. Prepublication data sharing. *Nature* 461:168–170. [doi:10.1038/461168a]
- Wicherts, J. M., D. Borsboom, J. Kats, and D. Molenaar. 2006. The poor availability of psychological research data for reanalysis. *American Psychologist* 61:726–728. <<http://users.fmg.uva.nl/jwicherts/datasharing.pdf>>
- Wolins, L. 1962. Responsibility for raw data. *American Psychologist* 17:657–658.

John H. Porter  
University of Virginia  
Charlottesville, VA

---