Chapter 2

History and Organization of the Ecotrends Project

C.M. Laney, D.P.C. Peters, and K.S. Baker

Cross-site synthesis initiatives offer important opportunities for learning. The internal organizations and histories of these projects are not always documented in detail, but their lessons can inform future projects or sites that would like to participate in larger projects (chapters 16 and 17). In this chapter, we describe the internal organization and timeline of the EcoTrends Project as background to the data and recommendations that follow in subsequent chapters.

The EcoTrends Project began in 2004 when two scientists (Debra Peters and Ariel Lugo) saw a need to synthesize, and make easily accessible, long-term datasets to compare continental-scale and national-level trends in ecological responses to changing environmental drivers (figure 2-1). Because Peters (of USDA, Agricultural Research Service [ARS]) and Lugo (of USDA, Forest Service) are employed by different Federal agencies with existing networks of sites and were actively involved in the Long Term Ecological Research (LTER) program, the EcoTrends project began as a multiagency collaboration, initially funded by ARS. The project's organizational structure expanded over the next 6 years to include many activities and dozens of individuals from six major groups.

Project Organization

Broad organizational structures and a well-defined set of objectives and communication processes were needed to make the project successful. These arrangements were a critical aspect of the project because of the data management differences between sites and agencies as well as the large variety and number of datasets. The six major groups (figure 2-2) each contributed to infrastructure and produced new knowledge and data products (table 2-1):

1. The EcoTrends Project Office (EPO) in Las Cruces, NM, consisted of a director (scientist) (D. Peters), a project coordinator (C. Laney), a spatial analyst (J. Yao), and several graduate and undergraduate student assistants. The information manager of the Jornada Basin LTER (JRN) (K. Ramsey) assisted with designing, building, and maintaining the in-house information management system. The EPO provided overall direction and leadership for the project and worked closely with the other five entities to assemble, correct, and verify long-term data and metadata; to create the derived data products; to coordinate documentation of the derived datasets; and to make them publicly available via a website (http://www.ecotrends.info). ARS and JRN began funding work at EPO in 2004. National Science Foundation supplements to the JRN site provided support for the period 2006-2009.

2. The EcoTrends Editorial Committee (EEC) was formed in 2005 and consisted of a group of 12 scientists (authors of this book) with different expertise (including population ecology and biogeochemistry) and experience with different habitat types (such as lakes, urban, forests, grasslands, oceans) or system components (plants, animals, soils). Members of this committee sorted through the vast amounts of historic and ongoing data from all 50 sites and made decisions about the variables to be included and the content and organization of the book and the website.

3. The EcoTrends Technical Committee (ETC) was also formed in 2005 and consisted of a group of nine computer scientists and information managers drawn from the LTER Network Office, the National Center for Ecological Analysis and Synthesis (NCEAS), and the LTER information managers. Members of this committee provided advice on data and metadata best practices and functionality of the website. The members of this committee are the technical consultants for this book.

4. Participating site scientists, information managers, and technical staff were engaged in the project at various times and provided their datasets to EPO, verified data quality and quantity, and assisted EPO in creating corrected, derived datasets. They provided important insight into the needs of site personnel, issues with creating and comparing derived datasets, and the lessons learned while building their own information management systems and while coordinating data and information transfer with other sites.

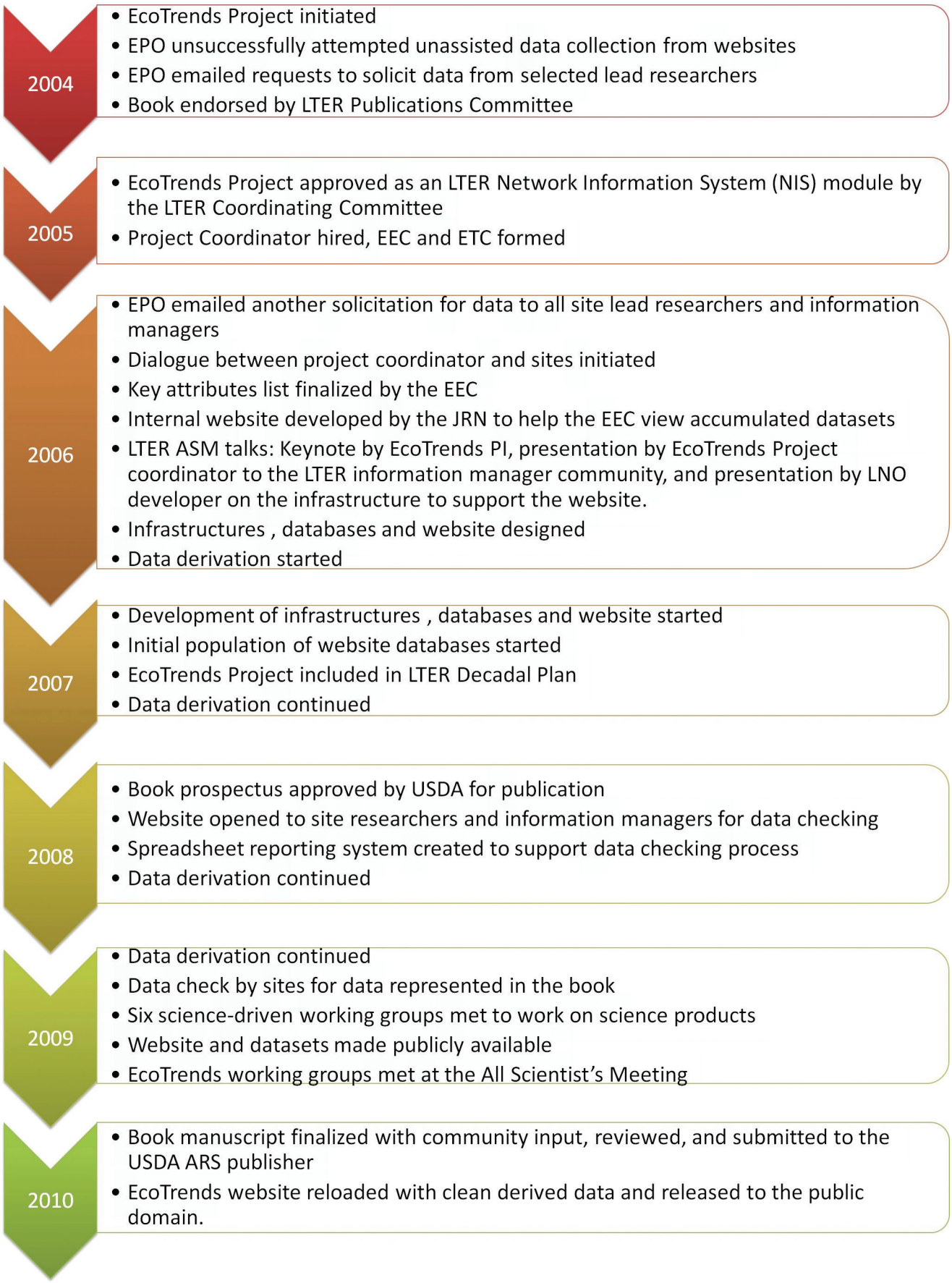

Figure 2-1. EcoTrends timeline from 2004 to 2010.

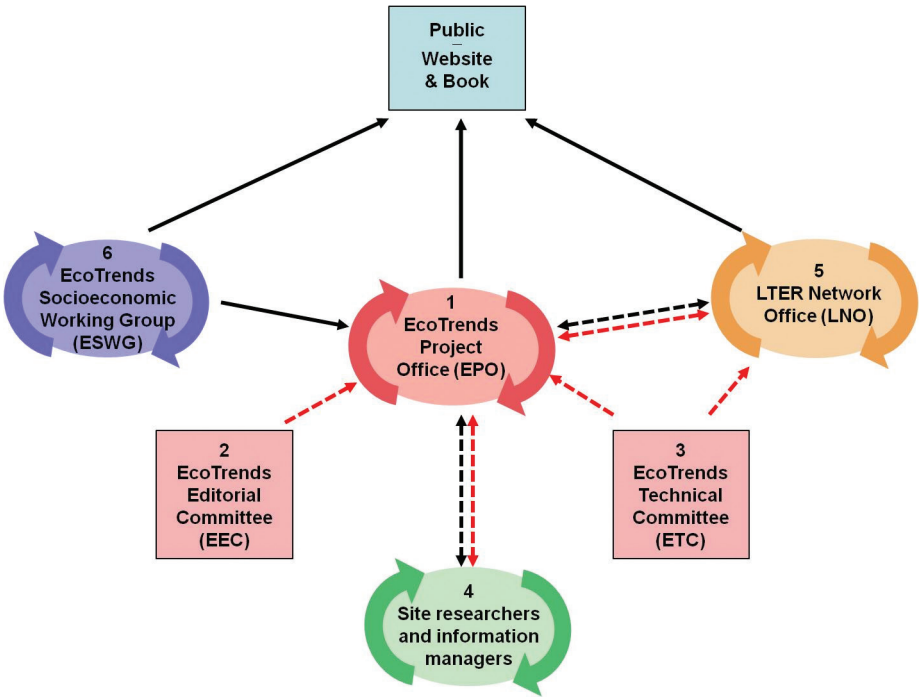

Figure 2-2. EcoTrends organizational arrangements and products. Each work arena is depicted by an ellipse with thick curved arrows that represent internal, dynamic information systems. The advisory committees are shown as rectangles. Straight arrows indicate interactions between the work arenas. Solid black arrows show dataset transfer. Dashed black arrows depict communications between arenas about data issues. Dashed red arrows depict flow of advice.

5. The LTER Network Office (LNO) formed parts of the EEC and ETC, helped design the EcoTrends website, developed routines to create derived dataset documentation and to support website functionality, and deployed the website from its local servers. LNO provided travel support for meetings of the EEC in 2006-08. National Science Foundation supplements to LNO supported work from 2006 to 2009.
6. The EcoTrends Socioeconomic Working Group (ESWG) was composed of one member of the EEC (J. Grove) and two LTER scientists (T. Gragson and C. Boone). This group used supplemental funding from the U.S. Department of Agriculture, Forest Service and National Science Foundation to New Mexico State University to compile historical census data for the participating sites (comprising about 1,000 counties and 32 variables) from several sources. This group also developed a complementary website, the LTER Socioeconomic Catalog (table 2-1), to make these data publicly accessible. A subset of these data were used in this book and are posted on the EcoTrends website.

Timeline

Gathering datasets took a substantial amount of time and effort by a large number of participants in all six groups. Dataset gathering began in 2004 when an undergraduate student from New Mexico State University was hired to find, download, and document long-term datasets (10 years or longer) from websites of research sites. However, this task was more substantial than anticipated. Few web pages provided tools to differentiate long-term datasets within large data stores. Some datasets were insufficiently documented or quality checked and verified for accuracy. Accordingly, the EPO was expanded in 2005 to include a project coordinator and a support position at JRN. ECC and ETC were formed to help assess the status of the data gathering effort and to solicit further contributions. In addition, the project was approved as an LTER Network Information System module (Brunt 1998, Baker et al. 2000) by the LTER Network governing body (the LTER Coordinating Committee), and the book was approved as an LTER publication by the LTER Publications Committee.

Prior and subsequent to the ECC's first meeting in 2006, email solicitations for datasets, without restriction on variable type or documentation level, were sent to the lead scientist at each site. At some sites, requests were handled by the lead scientist or a team of ecologists. At other sites, the request was transferred to the site information manager who often responded

Table 2-1. EcoTrends entities: knowledge, infrastructure, and data products[1]

EcoTrends entities	Knowledge products	Infrastructure products	Data products
EcoTrends Project Office (EPO)	Synthesis book	Local database, project-level data repository	Derived datasets with metadata
EcoTrends Editorial Committee (EEC)	Cross-site scientific publications; advice for website front end and the EPO	Interactions and support between site researchers	Selection of data products
EcoTrends Technical Committee (ETC)	Technical publications; advice for website back end	Interactions and support between site information managers, LNO, and EPO	
Site researchers, information managers	Scientific publications	Information systems, including data repositories and digital libraries	Original datasets with metadata
LTER Network Office (LNO)	Technical publications	Information system, data repository, and website	
EcoTrends Working Group (ESWG)	Cross-site scientific publications	LTER Socioeconomic Catalog database and website	Population and economy datasets

[1]A distinction is made here between knowledge products (such as scholarly works), infrastructure products (such as database or website development), and data products (such as data tables, metadata documents, and graphs) (Gibbons et al. 1994, Hine 2006).

by sending datasets or links to online datasets. Several hundred datasets were submitted that were then categorized by common variable (such as temperature, nitrogen deposition, or plant cover) and examined for consistency among sites by the ECC. Where critical datasets appeared to be missing, followup e-mail requests were sent to the site contacts to check the availability of the datasets, resulting in further submissions.

In addition to the directly submitted datasets, data from other organizations were downloaded from public websites (See table 2-2 for definitions of acronyms and Internet links). Climate and hydrological data were downloaded from the LTER Climate Database (Henshaw et al. 2006), the National Climate Data Center (NCDC), the National Oceanic and Atmospheric Administration (NOAA), and the United States Geological Survey (USGS). Atmospheric chemistry data were downloaded from the National Atmospheric Deposition Program (NADP). The ESWG coordinated the downloading and processing of human population and economy data from the InterUniversity Consortium for Political and Social Research and GeoLytics (http://www.geolytics.com/). A nearly complete working list of key variables and datasets was agreed upon at the ECC and ETC meeting in July 2006 and confirmed at the following meeting in February 2007.

From 2006 to 2008, solicitation of site-level datasets continued while computer programs in R (http://www.r-project.org) were written to process and graph the data. Throughout this period, EEC communicated frequently with EPO to review data progress and make recommendations on further work. In 2008, EPO asked the LTER community to review source and derived datasets online in the form of tables and graphs. Dataset review was divided into several stages. Sites were first asked to check the derived climate, biogeochemistry, and human population data and some months later to review the complete set, including biological data. Site personnel were asked to review and update their source data when necessary.

Dialogue among members over design issues progressed over several years of database and website design and implementation. At the EPO, a database, a data store, and a versioning repository system were developed to track the source data, manage the derivation processes, and document the derived datasets. A local website was developed at JRN to assist with database management, to allow EEC to review book graphics remotely, and to comment on the products and overall progress of the project. The design process for the EcoTrends website also began. A website designer was contracted, and the initial website design was sent to LNO for refinement and implementation. LNO designed and developed an automated system for harvesting each derived dataset and associated metadata into the databases underlying the website, using the EPO database and file naming structures, and for generating an Ecological Metadata Language (EML) documentation file for each derived data product. LNO also built the underlying website structure and tools necessary for data searching, browsing, viewing, and visualizing graphically.

In 2009-2010, EPO tested the usefulness of the derived data and website through six scientist-led working groups. These groups, each working with a different theme, explored how synthesis of EcoTrends-derived datasets could inform research. Each group also explored the EcoTrends data repository, downloaded useful data from the website, and analyzed these data in the context of other non-time-series data. This exercise resulted in valuable feedback about the usability of the website and the data it contains.

Near the end of 2009, EPO asked all participants to extensively check in detail the graphics presented in this book, the derived data, and the associated content on the EcoTrends website, providing another opportunity for community-level participation. Each chapter of this book was written by a small set of site participants and posted online for review by all site participants. An early version of the EcoTrends website was made available to the participants to explore datasets, provide recommendations on future website redesign, and comment on missing data types. Although sites had been asked several times over the past couple of years to check their data, this final check elicited further feedback from the community, likely stimulated by the immediacy of seeing their data and text in print.

Table 2-2. EcoTrends project-related organizations: acronyms or terms and Internet links

Acronym/term	Name	Link
EML	Ecological Metadata Language	http://knb.ecoinformatics.org/software/eml/
EPA	Environmental Protection Agency	http://www.epa.gov
FGDC	Federal Geographic Data Committee	http://www.fgdc.gov
GeoLytics	GeoLytics demographic data	http://www.geolytics.com
ICPSR	Inter-University Consortium for Political and Social Research	http://www.icpsr.umich.edu
LNO	LTER Network Office	http://lno.lternet.edu
LTER	Long Term Ecological Research Network	http://www.lternet.edu
EcoTrends Socioeconomic Catalog	EcoTrends Socioeconomic Catalog	http://coweeta.uga.edu/trends/
Metacat	Ecoinformatics Metadata Catalog	http://knb.ecoinformatics.org
NADP	National Atmospheric Deposition Program	http://nadp.sws.uiuc.edu
NCDC	National Climatic Data Center	http://www.ncdc.noaa.gov
NOAA	National Oceanic and Atmospheric Administration	http://www.noaa.gov
R	R project for statistical computing	http://www.r-project.org
ARS	United States Department of Agriculture, Agricultural Research Service	http://www.ars.usda.gov
USFS	United States Department of Agriculture, Forest Service	http://www.fs.fed.us/

Contributions to Information Management

A set of formalized databases and communication systems were needed to address organizational and technological challenges of managing the hundreds of submitted and downloaded datasets (source datasets) within and between EPO and LNO. As projects of this size and scope are complex and relatively rare, advice on how best to proceed was needed from a broad community. ETC advised EPO and LNO on technical issues, data management practices, organizational mechanisms, and website development. Presentations made at various meetings engaged participants and elicited further input from the science and information management communities. EcoTrends information management also drew upon participants' past experiences with collaborative, cross-site research activities and existing network infrastructures, principally LTER.

Experience gained through data handling, web development, and technology committee and information management community discussions motivated the development of other LTER Network-level cyberinfrastructure projects, principally the Provenance-Aware Synthesis Tracking Architecture (PASTA) (Servilla et al. 2006, 2008). PASTA was conceived and prototyped to support the EcoTrends website, originally as the tool to automate harvesting of the derived data into a repository that was accessible to the website. The EcoTrends experience also contributed to further development of EML and of Metacat, a system developed by the Knowledge Network for Biocomplexity for cataloging EML documents.

Conclusions

The EcoTrends Project is a scientist-driven initiative that has, since 2004, drawn upon a large and diverse community of researchers, information managers, and computer scientists for advice and support. Interactive cycles of refinement were based on community feedback and lessons learned. Where possible, the project attempted to use and support further development of community data practices and metadata standards, while maintaining flexibility for datasets that did not fully meet these practices or standards. This approach facilitated an evolving trend toward data sharing and synthesis. Lessons learned throughout the process (chapters 16 and 17) will inform future multiagency, cross-site, multidisciplinary projects.

References

Baker, K.S., B.J. Benson, D.L. Henshaw, et al. 2000. Evolution of a multisite network information system: the LTER information management paradigm. BioScience 50:963-978.

Brunt, J. 1998. The LTER network information system: A framework for ecological information management. *In* C. Aguirre-Bravo and C.R. Franco, eds., North American Science Symposium: Toward a United Framework for Inventorying and Monitoring Forest Ecosystem Resources; 2-6 Nov 1998, Guadalajara, Mexico, pp. 435-440. Proceedings RMRS-P-12.

Gibbons, M., C. Limoges, H. Nowotny, et al. 1994. The New Production of Knowledge: the Dynamics of Science and Research in Contemporary Societies. Sage Publications, London.

Henshaw, D.L., W.M. Sheldon, S.M. Remillard, et al. 2006. CLIMDB/HYDRODB: a web harvester and data warehouse approach to building a cross-site climate and hydrology database. *In* Proceedings of the 7th International Conference on Hydroscience and Engineering (ICHE-2006); Philadelphia, PA; not paged. Drexel University College of Engineering, Philadelphia. Online: http://hdl.handle.net/1860/1434; Available through iDEA: Drexel E-repository and Archives.

Hine, C.M. 2006. New Infrastructures for Knowledge Production: Understanding E-Science. Information Science Publishing, London.

Servilla, M., J. Brunt, I. San Gil, et al. 2006. PASTA: a network-level architecture design for generating synthetic data products in the LTER Network. LTER DataBits, Fall 2006. http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/06fall/.

Servilla, M., D. Costa, C. Laney, et al. 2008. The EcoTrends web portal: an architecture for data discovery and exploration. Proceedings of the Environmental Information Management Conference 2008, pp 139-144.

USDA
United States Department of Agriculture

Agricultural Research Service

Technical Bulletin Number 1931

September 2013

Long-Term Trends in Ecological Systems:

A Basis for Understanding Responses to Global Change

Editors:

Debra P.C. Peters, Christine M. Laney, Ariel E. Lugo, Scott L. Collins, Charles T. Driscoll, Peter M. Groffman, J. Morgan Grove, Alan K. Knapp, Timothy K. Kratz, Mark D. Ohman, Robert B. Waide, and Jin Yao