Managing Data from Multiple Disciplines, Scales, and Sites To Support Synthesis and Modeling

R. J. Olson, J. M. Briggs,† J. H. Porter,‡ G. R. Mah,¶ and S. G. Stafford§*

The synthesis and modeling of ecological processes at multiple spatial and temporal scales involves bringing together and sharing data from numerous sources. This article describes a data and information system model that facilitates assembling, managing, and sharing diverse data from multiple disciplines, scales, and sites to support integrated ecological studies. Cross-site scientific-domain working groups coordinate the development of data associated with their particular scientific working group, including decisions about data requirements, data to be compiled, data formats, derived data products, and schedules across the sites. The Web-based data and information system consists of nodes for each working group plus a central node that provides data access, project information, data query, and other functionality. The approach incorporates scientists and computer experts in the working groups and provides incentives for individuals to submit documented data to the data and information system. Published by Elsevier Science Inc.

INTRODUCTION

The synthesis and modeling of ecological processes at multiple spatial and temporal scales involves bringing together and sharing data from numerous sources. An example of such an activity is the BigFoot project that will bring together investigators, ideas, and data for the purpose of locally validating the National Aeronautics and Space Administration Earth Observation System-era global data sets (Running et al., 1999, this issue). This coordinated exercise proposes to study the effect of scaling from a fine grain to a coarse grain on estimates of important biosphere variables from a range of biome conditions. Using standardized methods that incorporate extensive ground data sets, ecosystem models, and remotely sensed imagery, sites will develop local maps of land cover class (LCC), leaf area index (LAI), and aboveground net primary productivity (NPP). This article describes a data and information system (DIS) model that would facilitate assembling, managing, and sharing diverse data from multiple disciplines, scales, and sites to support integrated ecological studies, such as the BigFoot project.

* Oak Ridge National Laboratory, Oak Ridge, Tennessee
† Kansas State University, Manhattan
‡ University of Virginia, Charlottesville
¶ U.S. Geologic Survey, Sioux Falls, South Dakota
§ Colorado State University, Fort Collins

Address correspondence to Richard J. Olson, Oak Ridge National Laboratory, P.O. Box 2008, MS 6407, Oak Ridge, TN 37831-6407. E-mail: rjo@ornl.gov

Received 15 June 1998; revised 16 October 1998.

REMOTE SENS. ENVIRON. 70:99–107 (1999)
Published by Elsevier Science Inc.
655 Avenue of the Americas, New York, NY 10010

0034-4257/99/$–see front matter
PII S0034-4257(99)00060-7

Integrated, Multisite Projects

Projects that entail the synthesis and modeling of ecological phenomena over regions require multidiscipline data from multiple study areas or sites. Such complex projects often require an integrated database that is comprised of historical data, newly collected field data, remotely sensed image data, and GIS (geographic information system) coverages. The challenge of combining diverse data for global change research has been studied by a National Research Council review committee (NRC, 1995). On the basis of six case studies, the NRC committee identified 18 barriers to combining diverse data and developed 10 keys to success. Only two of the ten keys dealt with technology while the others addressed cultural aspects of management, human behavior, and marketing associated with developing integrated data resources. A companion challenge of ensuring the availability of ecological data has been studied by the Ecological Society of America (Gross et al., 1995). This study identified a growing concern that long-term ecological data are lost after projects are completed because investigators move on to other interests and the data are rarely archived. Both data management concerns of combining diverse data and providing ready access and long-term archive to data are addressed as an integral part of the proposed data and information system.

One of the projects examined by NRC as an example of a large integrated ecological study is the First ISLSCP (International Satellite Land Surface Climatology Project) Field Experiment (FIFE) funded by NASA (NRC, 1995). The FIFE project collected intensive remote sensing and field data on a prairie site in Kansas in 1987 and 1989 to better understand fluxes between the land surface and the atmosphere and to develop associated remote-sensing methodologies. The project included 29 teams organized in six disciplinary groups. An extensive collection of FIFE data and metadata were compiled and distributed as a 5-volume CD-ROM set (Strebel et al., 1994a), and these data are currently available online from the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC) World Wide Web home page (http://www-eosdis.ornl.-gov). The DAAC is the long-term archive for the data. The FIFE information system was also used to demonstrate the implementation of a conceptual framework for scientific information systems (Strebel et al., 1994b).

LTER Network Information System

Another example of developing an integrated database is the efforts of the Long-Term Ecological Research (LTER) Network to facilitate intersite research by linking individual site information systems. In the LTER Network, individual sites routinely collect daily climate data and maintain the data in local information systems using a variety of formats. Climate data are almost universally collected in multisite studies and are typically the most commonly requested data to perform synthesis and modeling. To facilitate such synthesis, the LTER community gathered individual site temperature and precipitation data and created online monthly summaries for each site (Henshaw et al., 1998). While these summaries satisfied a need for access to monthly site climate data, there was no provisions for maintaining and updating these summaries or satisfying frequent requests for daily data; therefore, a strategy was developed to provide climatic data dynamically (Henshaw et al., 1998). Each site provides access to standardized daily climate files via an Internet address that points to the location of static files or dynamic scripts. Daily climate data are harvested automatically by a central site into a centralized database and applications programs produce two monthly distribution reports or formats from the daily climate database. The challenge of developing and implementing this strategy is reflected by the fact that the process has evolved over 10 years, starting with the establishment of standards for baseline meteorological measurements in 1986 (Greenland, 1986).

In addition, the LTER information managers have led a concerted effort to develop a LTER Network Information System (NIS) (1995 LTER Data Managers Workshop, http://lternet.edu/im/dm95report.htm) to support basic ecological research, both at the site and network levels. When fully implemented, the system will improve Internet connectivity among the research sites and transform a collection of individual sites and databases into a coherent network. It will be a distributed system so that site data and information can be maintained locally. The scaleable system will enable users to expand site-specific studies into intersite studies and to use prototype systems or modules, that are developed for small groups or studies, for larger groups and intersite use. Queries and browsing of data sets and metadata located at multiple sites will be essential characteristics for the system to function as an intersite research tool. The system could have the capacity for data input at the site level to accommodate the needs generated by specific research groups. Currently NIS includes an investigator list, bibliography (Chinn and Bledsoe, 1997), data set directory, climate database, and a prototype compilation of site characteristics, with information and data from all sites in each. Future plans are to expand this system to encompass all types of ecological data from the LTER sites.

An important issue in developing the LTER Network Information System is the consistent use of existing content standards for data and metadata at all sites, as well as establishment of new content standards where they do not exist. Standards will be a prerequisite for the distributed system's capability to present data and information in a consistent fashion, independent of their original format and location. Standardization also is helpful for supporting meaningful synthesis of data across multiple sites. However, network-wide format standards, which may simplify implementation of the system, are less critical than the content standards, since a uniform appearance could be achieved by using filter programs to create standard views from files in different formats as maintained by sites. Nevertheless, data and metadata format decisions will still affect a variety of tasks, such as development of system parts such as a distributed cross-site catalog. This will require consistency between sites, standardized keywords, development of the catalog query system (minimum metadata standards, structure/format/access methods standard within site, writing filters to produce desired displays, etc.). Platform independence will make it possible to use the system in the heterogeneous hardware and software environments in use at different sites.

Data and Information System Model Goals

The data and information system model for multisite projects incorporates concepts tested in the FIFE project and in the evolving LTER Network Information System design. The goals are to develop a multitiered information system that supports the project, facilitates maximum use of the data by the projects' investigators, creates a data archive, and provides data access to other researchers. The data and information system model incorporates data processing as an integral component of ongoing activities associated with the project, especially among science working groups and at individual sites of a multisite project.

Key features associated with the model include:

- data will be available to both internal project users and outside users in a timely manner;
- appropriate credit will be given to the data originators;
- data will undergo quality assurance and enhancements into formats easily usable by scientists;
- metadata will be provided that are essential to the understanding and use of the data;
- data from dispersed sources will be compiled into standardized data sets or data products;
- the overall data requirements, formats, and schedules will be determined by science domains with oversight by the project leader;
- the data management component will require project resources shared between science domains and the project data and information system.

OVERVIEW

An information system model for multisite projects was developed by the authors at workshop in May 1996 associated with planning of a predecessor of the BigFoot project. The core component of the data and information system was perceived as a set of cross-site scientific-domain working groups (e.g., NPP, LAI, remote sensing, etc.). Each of the working groups would coordinate the development of data associated with their particular scientific working group, including decisions about data requirements, data to be compiled, data formats, derived data products, and schedules across the sites. This arrangement can be envisioned as a set of pyramids with the sites and their data at the bottom of each pyramid, data processing in the middle, and the final standardized data set(s) at the peak. The workshop participants stressed the need for providing incentives for individuals to prepare documented data sets. This model contrasts to traditional information system models for ecological research projects where information management may be relegated to a separate, isolated data group and the long-term maintenance of data may not be addressed. It is similar to the model discussed by Strebel at al. (1994b), with additional emphasis placed on responsibilities of scientific working groups and the long-term data distribution and archive center.

The overall data and information system consists of nodes for each scientific working group plus a central node (Fig. 1). The central node is the primary entrance to all of the data. It would include access control, project information, data query, data archive, and other functionality. The system is envisioned to be Web-based and accessible through one or more of the popular browsers using an html type interface to the data and information.

Scientific-domain working groups might include individuals collecting field measurements, those compiling remote sensing and spatial data, those developing models, and those aggregating and extrapolating spatial information. The group leader for each scientific-domain working group provides scientific and technical leadership. The group leader would play a critical role in the development of data sets and data products for each working group. They would be in a position to identify the critical data resource needs and quality assurance criteria because of their scientific expertise. Equally important, as a scientific leader in the project, they would have the prestige needed to encourage other participants to fully collaborate in the development of data sets, metadata, and data products.

Group leaders would insure that the multisite data for their node were either entered into the database or accessible in standardized formats within the overall project schedule. However, the leaders may not have the technical expertise in information management to fully implement the data compilation efforts. For this reason, one or more technically oriented partners for each leader will be required. The resulting partnership assures scientific credibility, computer efficiency, and timeliness with the leader providing guidance and clout and the partner providing expertise in database tools and networking. The project-level data activities would be performed by a project data staff comprised of the technical partners for the groups and a leader for the project data and information system. In addition, the project leader provides overall direction and coordination among the groups and sites.

The data and information system provides access to the complete, combined, consistent data at each node (some nodes may be located physically together). There may be links to other data archives that would provide access to project or related data. Access may be limited to data originators during the active phases of data compilation and analysis. However, as data sets become more mature, they become publicly accessible through the project data and information system. Finally, they are moved to a long-term archive and distribution center and are advertised through master data directories, such as the Global Change Master Directory (GCMD).

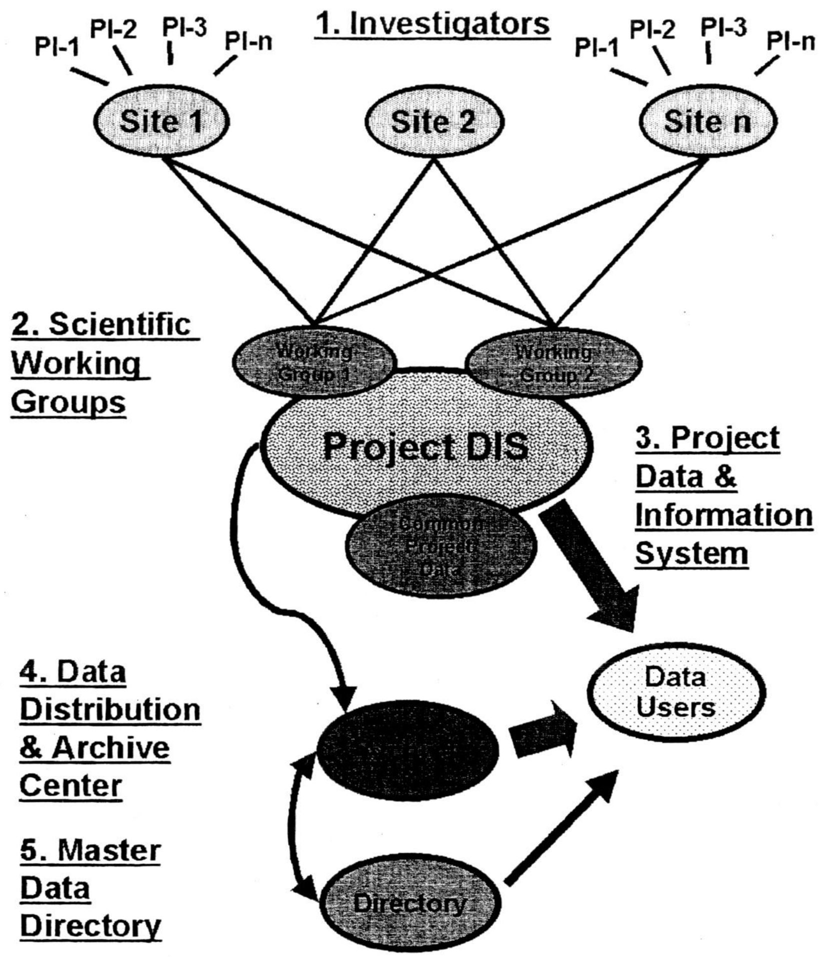

Figure 1. Information system model for managing and sharing data from multiple sites and disciplines to support ecological synthesis and modeling.

GENERIC ISSUES

Incentives for Compiling Data Sets

A key to compiling well-documented data sets and creating an integrated database is to encourage the cooperation of those who produce the data by providing incentives and rewards for database development (Olson et al., 1996). Some of the incentives that may be considered in developing the data and information system are summarized in Table 1. The process can also be limited by time and funds, authorship issues, insufficient global perspective (e.g., investigators may not see the need for recording latitude and longitude for a single site), a lack of standards and guidelines along with associated training, or the use of standards that do not fit complex ecological data and field situations.

Data Release Policy

Since October 1990, the LTER network has had guidelines (see Porter and Callahan, 1994) to promote making research data available to other researchers. To implement these guidelines, each LTER site has developed its own data management policy in consultation with key investigators and higher administrative units. In 1998 the guidelines were incorporated into a network-wide information management policy (Table 2). Table 2 provides guidelines and rationale, but each site is expected to defend its own implementation through the site and peer-review process. The LTER network has developed additional recommendations to alleviate concerns about the possible unethical use of data sets resulting from this policy; see http://lter.edu for more details.

Credits

To encourage individuals to compile data, mechanisms must be developed to give credit to the individuals associated with data compilation, especially in multicomponent projects. Credit will be conveyed by citations that indicate the individuals responsible for the data in the data set. The preferred citation, in a format similar to journal references, can be included in the metadata. Because data sets are sometimes combined for synthesis or broader modeling applications, a policy must be developed for citing data from multiple contributors. The scheme would be analogous to citing author(s) of chap-

Table 1. Incentives To Encourage the Development of Multiplesite Datasets

Goal	*Incentive*
Increase cooperation between scientists and data managers	Involve data managers in project planning and science
Promote data flow from scientists to data managers	Produce rewarding data products with deadline as joint effort between scientists and data managers, e.g., CD-ROM of data
Increased completeness and timeliness of data products	Allocate sufficient resources for data activities by investigators and data managers
Increase usefulness of the data	Require documentation as an essential and useful component of the project
Increase usefulness of the data	Adopt metadata standards (e.g., Michener et al., 1997)
Encourage scientists to compile data	Develop citation policy to give credit to data compilers, especially for datasets produced from multiple contributors
Protect scientist's exclusive right for initial publication of data analysis	Develop data release policy (e.g., see LTER data policy)

ters in a book, editor(s) of a book, and editor(s) of a series of related volumes. The following scheme is proposed for use by a multidiscipline project. If a article or report uses data from a specific data set, then the data originator(s) will be cited similar to a book chapter. If a article used data processed by a science team at the thematic or site level, then the leader of the science-working group would be cited similar to a book editor. Finally, if the entire project database was used, then an acronym can be used to represent the database and used to reflect contributions of all participants similar to a book series editor.

Data Enhancement

Even when well-documented data sets are available, they often must be augmented, or enhanced, before they can be used in new applications. Enhancing the data, as used in the context of this article, refers to data processing that is often required to perform integrated analyses and modeling within a multicomponent project. An enhanced data set is a data product derived from an initial data set and which is made more complete and internally consistent. The enhancement process addresses the practical aspects of incomplete data sets associated with fieldwork (e.g., problems associated with broken instruments, miscalibrated sensors, observer errors, and unusual field events). Missing values or outliers may be replaced with estimated values with appropriate flags and documentation of the estimation process. Calibration factors may be adjusted on the basis of an analysis of the project data. Data may be aggregated or extrapolated to create data sets with appropriate spatial and temporal characteristics. Associated site data—such as data related to soils, topography, and climate—may be acquired from other existing sources as needed by the models and become part of the enhanced data set.

Data Distribution

The Internet provides a powerful mechanism for facilitating the flow of information to a wide variety of users. However, this technology does not automatically solve the concerns about documenting the data, integrating disperse data, or creating a long-term data archive. Although the technology for creating individual Web home pages is widely available, the functionality of the sites varies, depending on the background and resources of the site administrators. The basic functions of a Web site and associated data center that distributes data should

Table 2. Data Access Policy for the LTER Network (November 1998)

1. There are two types of data: Type I (data that ae freely available within 2–3 years) with minimum restrictions, Type II [exceptional data sets that are avalable 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 data site 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

include distributing data on multiple media, allowing data browsing even by users with limited Internet capabilities, maintaining and distributing metadata, offering a data-search-and-order capability, providing data security, and providing a data archive. During the active phase of the project, all of the individuals, teams, and central information system may have active Web sites with links to each other. Web pages for individuals and teams may have limited functionality and accessibility while the central Web page has full functionality.

While the Internet is especially valuable for data exchange during the active phase of a project, CD-ROMs provide a valuable way to organize and distribute the collection of project data at the completion of the project. In fact, creating the CD-ROM package of data sets and metadata can provide motivation for investigators to complete individual data sets, especially the metadata, to be included on the CD-ROM. The CD-ROM is a static representation of the data, data products, and imagery. As a consequence, provisions for correcting errors, updating data sets, and adding new data and data products should be anticipated, such as maintaining lists of all individuals receiving CD-ROMs.

SYSTEM DESCRIPTION

Compiling multidisciplinary or multisite data to support a synthesis project will be accomplished by a data maturation process (Strebel et al., 1994b) in which data flow from *investigators* that collect and analyze field measurements through *working groups* that collate and standardize data to the *project level* for synthesis studies and, finally, into a long-term *data distribution and archive center* that provides data to secondary users. In addition, descriptions of the data set may be submitted to a *master data directory* with search functionality. Although the functions can overlap or several functions can be performed at one level, the general roles and responsibilities of individuals at each of the five levels are described below.

Investigator

Traditionally investigators collect and analyze their own data with little need for standardized data sets or written metadata. In a synthesis project, investigators assume responsibilities for collecting, inputting, performing quality control (QC) and quality assurance (QA), processing, and documenting their data following working group (next subsection) and project (subsection after next) guidelines and schedules for managing their data. Data preparation to meet project synthesis needs can be costly, often beyond the investigators' interests and resources; therefore, it is essential not only to provide guidance and resources but also to ensure that individuals receive credit and that incentives are given, such as co-authorship of articles that use the data from many sites. After review and authorization by the investigator, data would be made available to the appropriate project working group.

Working Group

The scientific-domain working groups have needs to standardize and collate data collected at the multiple sites to allow for combined synthesis and modeling. Often these activities are tied to running models and publishing articles by groups of scientists—this incentive is key to developing complete and consistent data sets for the project. Working groups would coordinate the development of data associated with their respective domains, including issues of content, consistency, and completeness. At the beginning of the project working group leaders, in partnership with their technical support and guidance from the project DIS (next subsection), would establish data requirements, data content standards, data collection methods, processing algorithms, data formats, derived data products, and data processing schedules.

Working groups would be responsible for performing QA checks and metadata reviews, collating the clean and documented data sets from the investigators, and possibly processing the collated data with standard algorithms. The QA checks include plots, statistics, and sorted lists; a review of logical combinations of variables; standardization of dates, times, coordinates, units, codes; and consistent data organization with consistent naming conventions. Comparing and combining data from multiple sites can often facilitate QA checks that otherwise are not normally performed on data from individual sites. If needed, the working group would develop ways for filling in missing or problem data. Developing the data sets and data products for the scientific-domain working group may be done locally (at the same location, using local equipment or central resources) or remotely (with the working group leader at different locations). The working groups may not have the interest or resources for maintaining and distributing the data sets after their synthesis studies are completed. Therefore, it is important for the working groups to anticipate the need to make their documented data available to a central node for synthesis and modeling within the project and for eventual distribution to users outside of the project.

Project DIS

The project data and information system (DIS) makes data from the scientific-domain working groups available for project-level synthesis and modeling. Data activities at the project level include providing coordination between the working groups so that the data from the working groups will fit together at the project level and implementing the overall system. The project leader and the DIS working group would oversee information re-

Table 3. Typical Primary and Secondary Data Processing Roles Performed by a Project Data and Information System and Types of Data and Information Processed at the Project Level

Primary Roles	*Secondary Roles*	*Project Information*
Establish data and metadata standards	Gather historical and background data	Experimental data and metadata
Prepare documented data sets	Distribute copies of large off-line data sets	Inventory of data needs and available data
Review metadata, data QA, consistency, and completeness	Process meteorological and other auxiliary site data	Geographic information system based data
Provide data processing support	Process satellite/aircraft remote sensing	Remote sensing satellite data
Develop project data and information system (Web-based)	Provide data analysis tools and support	Flight logs, videos, site photos, aerial photos, and maps
Manage the project data: backup, access control, etc.	Prepare datasets for public access	Project data policies, guidelines, and standards
Distribute project information and data as requested	Transfer data to a long-term data archive center	Contact list of individuals associated with project
		Annotated bibliography of key papers

sources and data products from the individual working groups for overall project needs. To expedite the flow of data during the later stages of the project, it is vital to involve the DIS group in all aspects of the project from the beginning.

The DIS group develops and maintains a central Web home page to distribute information and data to investigators within the project. It has primary responsibility for the central node and linking the other nodes together. The DIS staff may provide data processing support for project investigators, distribute project information, and work closely with investigators to prepare accurate and completely documented data sets. The DIS staff may include individuals associated with specific working groups, that is, technical partners with dual membership in the DIS. The primary and secondary roles of the project DIS are described in Table 3. Table 3 also lists types of project data and information that may be included in the DIS.

In addition to data and data products, the DIS group may write programs for loading data, applications software, and models. Some of these programs may be kept internally as tools for managing data or maintaining programs used to generate data sets. If these programs are appropriate to release with the data sets as companion files or separate files, then metadata is prepared and software submitted to the distribution and archive center.

The project DIS generally does not support data distribution to a broader set of users or commit to long-term archiving of the data after completion of the project. As data sets are completed or at the end of the project, clean data sets and metadata are submitted to a distribution and archive center for distribution to the general public.

Distribution and Archive Center

The long-term distribution and archive center maintains and distributes data and provides support to data users. The center works with the project members to identify how best to support the project and to establish a schedule for transferring data. Data sets are transferred to the center as the project finishes or as individual data sets are completed. Data sets are reviewed for consistency and completeness and entered into the center's information management system. The center must acquire financial and institutional support to provide for long-term archive of the data. The LTER Network Information System is evolving toward providing the distribution and archive functions for LTER data. The Oak Ridge National Laboratory DAAC for Biogeochemical Dynamics provides these functions for NASA and selected non-NASA field measurement data.

The distribution and archive center reviews metadata to ensure that they are accurate and complete and that they provide present and future users (those not directly involved or familiar with the project) adequate information to understand the nature of the data sets and to assess their appropriateness for purposes beyond the scope of the original project. The data are reviewed to ensure that they agree with the metadata and are consistent and complete. The center also extracts keywords and other metadata for use as search criteria that enable users with varying scientific backgrounds to locate and acquire data that meet their needs. Finally, the center acquires approval from data providers for final data products prior to public release and loads the data and metadata into the archive. Other data center functions include: providing data security with password protection and backups of data files; providing links to investigators, project science teams, and project data and information system Web pages; archiving data (multiple copies, long-term retention, migration to new media as required); notifying users of any errors, changes, or updates to the data; advertising and actively marketing available data resources among scientists and other users; maintaining standard and valid keywords; supporting a data search-and-order capability; providing user support; and supporting data needs of decision makers, educators, and the general public.

The center continues to maintain, distribute, and archive the data after the project finishes. The center supports the field project by providing a link between data providers and users. That is, the center generally may be able to answer many of the user questions, but if necessary the center could contact investigators for more information. In addition, the center may routinely generate statistics on users of the data, solicit information from the investigators on publications and new data products, solicit information from the investigators and user community about potential errors and updates, provide a mechanism for investigators to exchange ideas on proposed activities or improvements, incorporate new data and data updates, and communicate changes and updates to users.

Master Data Directory

Master data directories, or data catalogs, contain large collections of descriptions of data sets but do not work directly with or archive the data. The data directory contains pointers to the numerous locations that contain the data and therefore provides a mechanism for scientists to locate, acquire, and integrate data from various archives. Generally, directory-level metadata are very broad in their coverage so that they assist a scientist in learning what data are available for a particular area. The metadata allow a user to determine whether a data set is interesting enough that additional specific information should be sought. Directory-level metadata can usually be used to reject a particular data set as not appropriate; however, to fully determine whether the data set located in the directory is of real interest to the scientist, more information about that data set is required at another level of the metadata hierarchy. To acquire the actual data set described in the master data directory, the user must access the center or individual that maintains the data set.

The NASA-funded Global Change Master Directory (GCMD) is an example of a master data directory. A long-term goal of this effort is to provide science users with the ability to find and view information about science data regardless of what data system actually holds the metadata. The directory metadata interoperability has been addressed through two mechanisms: the development of a Directory Interchange Format (DIF) and the development of a DIF-based master directory that contains data across agencies, disciplines, and international boundaries. The DIF contains about 35 descriptive fields with controlled lists of keywords for many of the fields. The GCMD, including a description of the DIF structure, can be accessed at http://gcmd.gsfc.nasa.gov/.

SUMMARY

The data and information system model described in this article facilitates assembling, managing, and sharing diverse data from multiple disciplines, scales, and sites to support integrated ecological studies. It allows bringing together and sharing data from numerous sources for the synthesis and modeling of ecological processes at multiple spatial and temporal scales. The core component of the system is a set of cross-site scientific-domain working groups. Each of the working groups would coordinate the development of data associated with their particular scientific working group, including decisions about data requirements, data to be compiled, data formats, derived data products, and schedules across the sites. Therefore, the proposed approach incorporates scientists in the working groups and provides incentives for individuals to submit documented data. The overall data and information system consists of nodes for each working group plus a central node. The central node is the primary entrance to all of the data and includes access control, project information, data query, data archive, and other functionality. The system is envisioned to be Web-based and accessible through one or more of the popular browsers using an html-type interface to the data and information.

Ideas and experiences from many individuals are incorporated in this article. The authors especially acknowledge the continuing dialog on developing data and information systems within the LTER network of data managers, the DAAC system development, and data centers at ORNL. Comments from three anonymous reviewers are also greatly appreciated. This research was sponsored by the National Aeronautics and Space Administration. ORNL Environmental Sciences Division Pub. No. 4900. ORNL is managed by Lockheed Martin Energy Research Corporation for U.S. Department of Energy under contract DE-AC05-960R22464. The U.S. government retains a non-exclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

REFERENCES

Chinn, H., and Bledsoe, C. (1997), Internet access to ecological information: the U.S. LTER All-Site Bibliography Project. *BioScience* 47:50–57.

Greenland, D. (1986), Standardized meteorological measurements for long-term ecological research sites. *Bull. Ecol. Soc. Am.* 67:275–277.

Gross, K. L., Pake, C. E., and the FLED Committee Members (1995), *Future of Long-Term Ecological Data (FLED), Vol. 1: Text of the Report*, Ecological Society of America, Washington, DC.

Henshaw, D. L., Stubbs, M., Benson, B. J., Baker, K., Blodgett, D., and Porter, J. H. (1998), Climate database project: a strategy for improving information access across research sites. In *Data and Information Management in the Ecological Sciences: A Resource Guide*, (W. K. Michner, J. P. Porter, and S. G. Stafford, Eds.), LTER Network Office, University of New Mexico, Albuquerque, NM, pp. 123–127.

Michener, W. K., Brunt, J. W., Helly, J., Kirchner, T. B., and Stafford, S. G. (1997), Non-geospatial metadata for the ecological sciences. *Ecol. Appl.* 7:330–342.

National Research Council (NRC) (1995), *Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data*, National Academy Press, Washington, DC.

Olson, R. J., Voorhees, L. D., Field, J. M., and Gentry, M J. (1996), Packaging and distributing ecological data from multisite studies. In *Proceedings of the Eco-Informa '96, Global Networks for Environmental Information*, Environmental Research Institute of Michigan, Ann Arbor, pp. 93–102.

Porter, J. H., and Callahan, J. T. (1994), Circumventing a dilemma: historical approaches to data sharing in ecological research. In *Environmental Information Management and Analysis: Ecosystem to Global Scales* (W. K. Michener, J. W. Brunt, and S. G. Stafford, Eds.), Taylor and Francis, Bristol, PA, pp. 193–203.

Running, S. W., Baldocchi, D., Turner, D., Gower, S. T., Bakwin, P., and Hibbard, K. (1999), A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sens. Environ.* 70:108–127.

Strebel, D. E., Landis, D. R., Huemmrich, K. F., and Meeson, B. W. (1994a), *Collected Data of the First ISLSCP Field Experiment, Vols. 1–5*, published on CD-ROM, National Aeronautics and Space Administration, Greenbelt, MD.

Strebel, D. E., Meeson, B. W., and Nelson, A. K. (1994b), Scientific information systems: a conceptual framework. In *Environmental Information Management and Analysis: Ecosystem to Global Scales* (W. K. Michener, J. W. Brunt, and S. G. Stafford, Eds.), Taylor and Francis, Bristol, PA, pp. 59–85.