

Organizational Informatics: Site Description Directories for Research Networks

Karen S. Baker
Scripps Institution of Oceanography, University of California San Diego
La Jolla, California USA 92093-0218 USA

and

James W. Brunt
Department of Biology University of New Mexico
Albuquerque, New Mexico 87131-1091 USA

and

David Blankman
Department of Biology, University of New Mexico
Albuquerque, New Mexico 87131-1091 USA

ABSTRACT

A site description directory plays a central role as a catalog for a network of research sites. Such a directory represents a keystone element in an information management system. A directory contributes to community communications both through documentation of member information and relationships as well as through design feedback elicited from participants in the ongoing process of developing the catalog system. Presentation of a description directory for networked research sites via web interfaces permits distributed, remote site data input and access. There is a dual challenge in creating an extensible directory design: first to capture relevant content and second to incorporate such a system within the work practice of the community represented to ensure its continued

evolution. We present here a directory designed for the Long-Term Ecological Research (LTER) Network community

Keywords: directory, metadata, organizational informatics, system design

INTRODUCTION

To describe, classify, and catalog are tasks fundamental to science. A site description directory extends the traditional catalogs of collections of objects (e.g., butterflies or rocks) and of data sets (e.g., temperature or biomass) to catalogs of systems such as research sites and networks. Although there is a growing understanding of the concept of data richness in the field of ecology today, there is less

Table 1: System and Sub-System Levels

Level	System	subsystem	Example	Organizational Principal
I. Cooperation	Individual Site Cross-site	Data1...Data N Data set1...Data set N Site1...Site N	Researcher A Site researchers A & B LTER sites A & B	Related data Related projects Related data sets & project
II. Federation	Discipline Domain Cross-Domain	Network1...NetworkN Partner1...Partner N Discipline1...Discipline N	LTER, NADP, OBFS OBFS, NEON Ecology, Earth Science	Related themes Related domain Related system

The International Institute of Informatics and Systemics
IIIS 7:355-360, 2002
Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics
14-18 July, 2002 Orlando FL

recognition that each research site and network is information-rich at an organization level as well.

Like the “experimental unit” in statistics, the “atomic” unit of ecological information management is the data set. A data set often can be presented as a table and its associated metadata. The content of data sets in ecosystem science ranges in complexity from measurements of daily temperature at a particular location to measures of diversity for a particular ecosystem. A table entity is an abstracted description, either by direct measurement or by derivation of some aspect of a physical entity or interaction of physical entities. A site or network is also an entity acting as a source of data or the subject of metadata. Much of the data in this directory constitutes the metadata describing research programs. In this work, “site” is used to describe a research team comprised of some combination of individuals united by a common study through an information and social structure. A site may be a member of a network or an association of sites.

The LTER directory design builds upon the structural similarity and significance of systems associated with both subsystems and larger systems (see Table 1). A multi-tier directory schema of metadata assumes sampling regions associated with a Site, a Site as a member of (related to) Networks, and Networks as related to Federations (Sheth and Larson, 1990).

A prototype **site description directory** for research sites has been designed for use as a module within the Long-Term Ecological Research Network Information System (NIS, Baker et al, 2000; Brunt 1999). The working model gathers and displays descriptive site data in addition to responding to user queries. It is a multi-level (network, site, subsite) iterative schema with attention to portability. The current research network **site description directory** is a two-tier implementation with a centralized relational database back-end and web-based user interface for data input, modification, display and comparison. The data model is relational with some object-relational aspects and with categories and themes identified by scientific participants themselves.

DESIGN ELEMENTS

A site description directory provides answers to questions such as

“What biomes do the sites represent?”
“What are the locations of the polar sites?”
“How large is the forest site?”

The needs met by a site description database include creating a repository of information that can be queried for a single site or across multiple sites, delivering easily accessible views of information in a common format and providing a mechanism for participant management of local information (add and/or modify). The choice of directory content material is important; it provides a common template that defines a site. Since local site definitions can be established independent of a network catalog, a site description directory can enhance organizational identity by making basic information about an association of sites available without detracting from local site autonomy.

Existing catalogs provide examples of working directory models. The initial LTER Network approach provides a list of links to each member’s web page where content presentation is independent (member in this context means the site organization, not each individual associated with the network). The National Atmospheric Deposition Program (NADP; <http://nadp.sws.uiuc.edu/nadpdata/>), with more than 200 network participants, and the Organization of Biological Field Stations (OBFS; <http://www.obfs.org/Members/StationList.html>), with more than 400 research field stations, represent a range of technology capabilities. Each met the challenge of diversity by compiling responses to an online form requesting organizational information that is maintained in a centralized location and so is ultimately queryable. Currently, in partnership with the LTER Network Office and the National Center for Ecological Analysis and Synthesis, OBFS is in the process of moving from a static to a dynamic presentation of information.

The structure identified as meeting both LTER directory needs and design criteria consists of three interrelated categories of member information: organization, personnel, and descriptive material (see Figure 1) organized into tables and including category or look-up lists. Information about

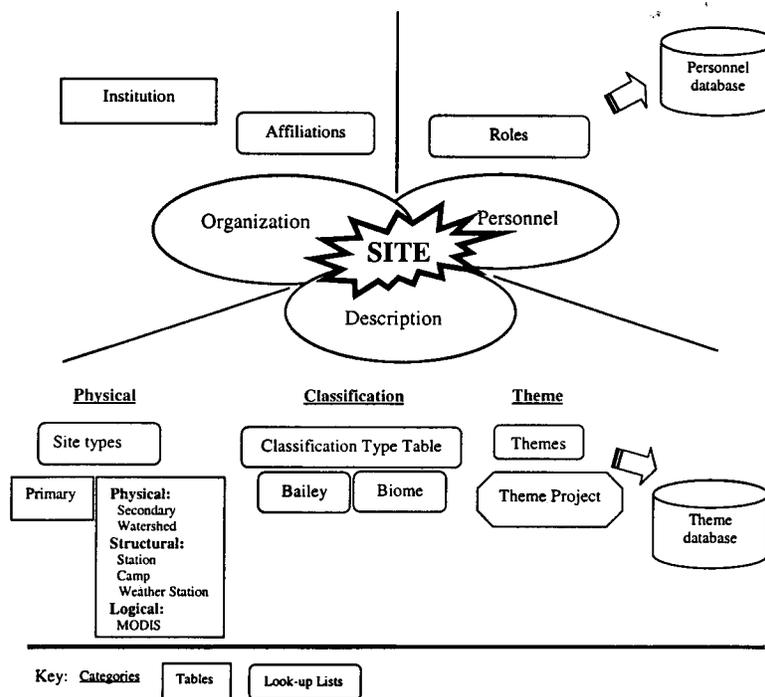


Figure 1. Site description directory components.

participating institutions such as their address and contact details along with their affiliations is presented in a directory as is personnel information including participant names, locations, email addresses and roles. Within the descriptive materials are subcategories of physical descriptors (e.g., latitude and longitude), classification tables (e.g., biome types) and theme lists (e.g., regionalization).

DESIGN APPROACH

The original design focused on simplicity. The data content represents a research site's general context. Discussions were held to engage and elicit the different participant views from within the organization so that the initial list of parameters could be shortened without losing critical information (or support). Information was eliminated that was either too detailed or too site specific in order ultimately to maximize site participation by minimizing time needed to complete forms. An early extension of the project was the addition of uniform resource locators (URL) in order to have available links to harvestable materials such as climate database files, site photos, and/or site maps.

Web based forms provide an interface for viewing

and comparing entries. The main view, or site view, is the root module. Further views may be topic specific such as climate, vegetation, regionalization, and soils. Categories are defined to use emergent classifications such as the Terrestrial Ecological Monitoring Stations (TEMS), a pilot project of the Global Terrestrial Observing System (GTOS). The Profiles of Ecologists (1992) contains a survey of the membership with a classification scheme for fields of ecological research and areas of expertise; the categories are empirical and include research themes as well as land types. Three variables are included for site classification description including the Bailey ecoregion types (Bailey, 1998) included in the TEMS/GTOS database. The biome group (i.e. tundra) and region classification (city, state, upper Bavaria regionalization) schema definitions require further definition as does the physical classification system developed through multi-site discussions.

Multiple views are available from the descriptive component information: single member information (home view), selected information from multiple sites (element view) or participant information on selected topics (theme view). The ability to compare parameters across sites demonstrates the value of a database approach.

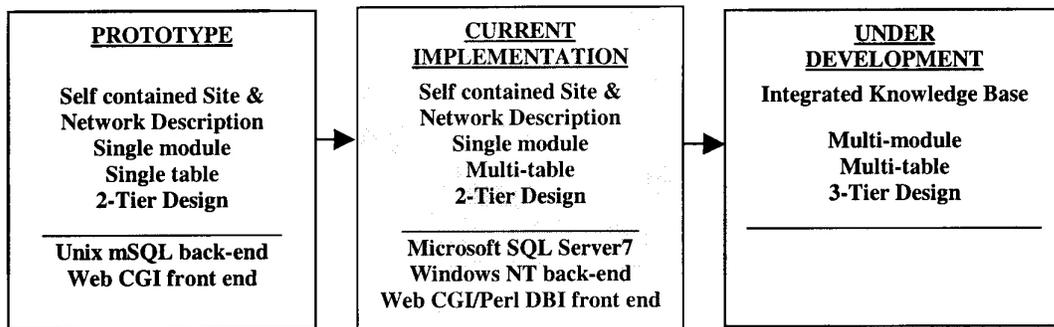


Figure 2. Three stage design development.

A set of web-based forms is used for adding, modifying and deleting from the directory. The entry forms are divided by view (theme): Site, Climate, Vegetation, Regionalization, and Soils. The modular entry forms approach breaks the information down into more easily digested units. A new subsite or theme may be added through the addition of a new table. Currently, the input tables include the site table with two forms or categories (description and URL), subsite tables with three categories (location, class, abiotic parameters), and research theme tables with a single category. Data entry is possible by designated site personnel with the site username/password login capability.

Design/Implementation Lineage

The design lineage of this project is outlined in Figure 2. The initial prototype design was driven by the design principle of simplicity, sacrificing flexibility, extensibility and modularity. The current implementation adds flexibility and some modularity at the expense of simplicity in order to broaden and deepen content. The current design is not easily portable with MS SQL Server 7 specific SQL calls along with mixed-case field and table names. While SQL Server (and Windows in general) is not case sensitive, the common, high-level programming language PERL (www.perl.com) is. Thus modification is required if the back-end is ported to Oracle or to non-windows platform.

The current implementation is also not interoperable with existing LTER database modules, specifically the LTER personnel database. For example, the current LTER personnel database uses an alphanumeric key to uniquely identify each person entered in the database. The directory database, however, uses a numeric key to identify people.

Current/Future Development

The design principles guiding current development efforts are:

Interoperability:

- Outward focus: Facilitate exchange with other network systems by using current and emerging metadata standards such as the Ecological Metadata Language (EML) for environmental data (<http://www.informatics.org>) and the Federal Geographic Data Committee (FGDC) standards for geospatial data (<http://www.fgdc.gov>).
- Inward focus: Develop XML web services to enable two-way communication between the centralized database and individual site databases that store similar information.
- Technical focus: Develop the ability to use different types of data stores (RDBMS, LDAP, XML Native) and exchange data with other metadata management systems such as the Metacat as part of the Knowledge Network for Biocomplexity (Jones et al, 2001).

Extensibility, flexibility and modularity:

- Extensibility: Develop the ability to use this information system for networks other than LTER or for sites within LTER that are also part of other networks.
- Flexibility: Develop the ability to add classification or descriptor systems from participant feedback without having to redesign the underlying database schema. This is being accomplished by i) Further normalization of the database schema; ii) Development of hierarchical representations within an RDBMS schema; Increased abstraction of the data model.

- **Modularity:** The original directory database is now a module in an integrated knowledge base that includes personnel, bibliography, document archive, event and meeting tracking, and grant tracking modules.

Scalability and portability:

- **Portability:** A three-tier design will be used to separate the user interface from the back-end data storage system. The system will be designed as a web service. The current plan is to use a J2EE compliant middle-tier using java data-typing, generic (platform independent) SQL calls with business rules defined in XML documents (Muench, 2001; OTN, 2002). This will allow the back-end storage system to be moved to a different database vendor or operating system without the need to modify the code on the user interface.
- **Scalability:** A three-tier system using lightweight data access objects makes the least demand on network and system resources. The platform independence discussed earlier also makes scalability possible.

Redesign is activated through testing of the site directory prototype in order to gain insights from user feedback. User feedback to date suggests addition of description information such as site directions, a biodiversity theme and capture of update times in addition to development history. Note, each addition preserves a bit of an organization's history.

Testing, a major task often neglected in the rush from design to production, requires time and support yet is essential to guarantee database module robustness. At best, practitioners participate in testing to ensure redesign usability and utility. To the extent that such a project incorporates local participation through identification of common information and classifications, organizational definition is enhanced through design anchored in practice.

CONCLUSION

Organizations, associations and partnership present the challenge of presenting member information that is manageable and accessible. A **site description directory** provides a mechanism to gather information about sites within a research network in

a common format. The design permits sites to update records, to add tables as needed to an extensible schema, and to store site URLs. The importance of a directory effort is that its impact is immediate for an organization, providing infrastructure cohesion and presenting metadata for query.

A directory enhances group communications. There is an increasing emphasis on partnership science as an approach to conducting science research such as with recent discussions regarding a National Ecological Observatory Network (NEON, <http://www.sdsc.edu/NEON>). Such associations bring requirements for new methods to manage organizational information. We are at a point where both the maturity of communities and of technological tools supports innovation in establishment of organizational infrastructure. Focus on infrastructure development activates communications (Hutchins, 1995; Kies et al; Robbins, 1995). Such an effort provides a method to stimulate system self-definition and to explore cognitive ecosystem concepts (Tomlinson et al, 1998; Schatz, 1993; Star and Rhuleder, 1994) where the term 'cognitive ecosystem' is used to describe the interdependence of a community's distributed knowledge and its social process.

Given the LTER organizational paradigm of participatory governance in addition to a full and synergistic partnership between science and information management (Stafford et al, 1994), the LTER network can serve as a valuable test-bed for considering methods to optimize communication and management through collections of organizational information. The LTER provides an opportunity to consider how a directory design can incorporate elements of social design to enhance the availability of organizational information.

A group of associated research sites comprise a network or a federation when committed to common goals or characterized by interdependence. As routine adoption and predictability are replaced by speed and innovation, associations require dynamic, integrative systems capable of evolving with the discipline. Computer technology, first seen as a provider of powerful computational engines, has evolved to provide methods for data organization and delivery in addition to becoming a potential factor in social change. The emerging concept of **organizational informatics** with a focus on information and communication infrastructure

provides an enabling vision in community efforts to transform distributed elements into an integrated system.

ACKNOWLEDGEMENT

Recognition and thanks to the collaborative design and implementation by Weimin Li. NSF Grants OPP-96-32763, DBI-01-11544, and EIA-01-31958 (ksb) support this work as well as DEB-96-34135 and DEB-99-80154 (jwb, db). Emery Boose, Nicole Kaplan, and Susan Stafford provided helpful comments.

REFERENCES

- Bailey, R.G., 1998. Ecoregions. New York, Springer-Verlag.
- Baker, K.S., B. Benson, D.L.Henshaw, D. Blodgett, J.Porter, S.G.Stafford. 2000. Evolution of a Multi-Site Network Information System: the LTER Information Management Paradigm, *BioScience* 50(11): 963-978,2000.
- Brunt J.W., 1999. The LTER network information system: a framework for ecological information management. Pages 435-440 in C. Aguirre-Bravo, C.R. Franco (eds) *North American Science Symposium: Toward a Unified Framework for Inventorying and Monitoring Forest Ecosystem Resources*; 2-6 Nov 1998; Guadalajara, Mexico. Fort Collins (CO): US Dept of Agriculture, Forest Service, Rocky Mountain Research Station. Proceedings RMRS-P-12.
- ESA Profiles, 1992. Holland, M.M, D.M. Lawrence, D.J. Morrin, C. Hunsaker, D. Inouye, A. Janetos, H.R. Pulliam, W. Robertson, J. Wilson (eds). *Profiles of ecologists: results of a survey of the membership of the Ecological Society of America*, Washington, DC.
- Hutchins, E., 1995. *Cognition in the Wild*, The MIT Press, Cambridge, MA
- Jones, M.B., C.Berkley, J. Bojilova, and M. Schildhauer, 2001. *Managing Scientific Metadata*. IEEE Internet Computing Sep-Oct 2001. (<http://computer.org/internet>)
- Kies J.K., R.C. Williges, M.B. Rosson, 1998. Coordinating computer-supported cooperative work: a review of research issues and strategies. *Journal of the American Society for Information Science* 49: 776-791.
- Muench, S. and BC4J Development Team, 2001. *Simplifying J2EE and EJB Development with BC4J*. (http://otn.oracle.com/products/jdev/htdocs/j2ee_bc4j.html)
- Oracle Technology Network, 2002. *Querying XML in a Standard Way, on the emerging standards for querying XML: SQL/XML and XQuery*. (<http://technet.oracle.com/tech/xml/>)
- Robbins, R.J., 1995. Information infrastructure for the human genome project. *IEEE Engineering in Medicine and Biology Magazine* 14: 746-759.
- Schatz, B.R., 1993. Building an electronic community system. Pages 550-560 in Baecker R, ed. *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*. San Mateo, Calif.: Morgan Kaufmann Publishers, Inc.
- Sheth, A.P. and J.A. Larson. 1990. Federated database systems for managing distributed, heterogeneous and autonomous databases. *ACM Computing Surveys* 22: 183-236.
- Stafford, S.G, J.W. Brunt, and W.K.Michener, 1994. Integration of scientific information management and environmental research. Pages 3-19 in W.K.Michener, J.W. Brunt and S.G. Stafford (eds). *Environmental Information Management and Analysis: Ecosystem to Global Scales*.
- Star S.L., K. Ruhleder, 1994. Steps toward an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. Pages 253-264. *CSCW'94: Transcending Boundaries: Proceedings of the conference on Computer Supported Cooperative Work*, 22-26 October, Chapel Hill, NC. New York: ACM Press.
- Tomlinson, K.L., M.A. Spasser, J.A. Sanchez, and J.L. Schnase, 1998. Managing cognitive overload in the Flora of North America project. *Proceedings of the Thirty-First Hawaii International Conference on System Sciences (HICSS-31)*