Common Information Management Framework: in Practice

Eda C. Meléndez-Colom
Luquillo Long-Term Ecological Research Program, Institute for Tropical Ecosystem Studies,
Natural Science Faculty, University of Puerto Rico
San Juan, PR 00936-3682

and

Karen S. Baker
Palmer Long-Term Ecological Research Program, Scripps Institution of Oceanography
University of California, San Diego
La Jolla, CA 92093-0218

ABSTRACT

A common goal of information management systems (IMS) is to share information among its users and originators. These systems are usually implemented by project managers and sponsors. The design and implementation of an IMS in a research organization, prompted by directives from sponsors and mediated by the vision of information managers, reflects the activities performed by a research team (science investigators, students, and field and laboratory technicians). The Long-Term Ecological Research program is presented as an example of a data sharing community whose distinctive goal, in addition to sharing, is to preserve information for future generations. A Common Information Management Framework (CIMF) is the product of the LTER community interactions and the information tools to achieve their goals. At the same time, it provides an integrated platform to promote, facilitate, and guide the members of an LTER community in the management of the information they generate.

Keywords: management framework, community, information management, system design, LTER information system, data sharing and preservation

1. INTRODUCTION

To share and preserve information are the principal goals of developing common management systems within a spectrum of communities (IRRI, 2001; Water and Sewer Board of Montgomery, Alabama, 2001). Three system components can be identified: drivers that motivate the conception (design, definition, and implementation), infrastructure components that define data management, and a community that defines and uses this framework.

The Long-Term Ecological Research (LTER) program provides an example of a research community sharing data. The LTER consists of twenty-four independent research sites, each studying a specific ecosystem, linked into a network of sites supported by a Network Office. Research activities begin with field data collection and include data management. Depending on the type of physical environment where the studies are performed, the LTER sites collect data that range from plankton, fungal, and vertically integrated acoustic biomass data to climate, water and soil chemistry, animal, vegetation and population, and disturbance data such as hurricanes, fires, and lightning strike data. The requirement that each site designate an information manager ensures initiation of site data management, critical for long-term research but often untended for short-term projects.

An LTER information management (IM) system must balance responsiveness to immediate local research needs with a plan for long-term data storage. Within a research community, the needs are to facilitate short-term and long-term science, to enable analysis and synthesis as well as to capture and archive information for future generations (Bowser, 1986; Michener and Brunt, 2000).

The role of an LTER information manager is to design and develop a digital framework that supports and reflects the community's shifting research activities. We extend the term *common management framework* originally used for describing the centralized set of data, computational tools, and schemas for organizing information at an LTER site [11] to the term *Common Information Management Framework* (CIMF) which includes the drivers and the community that interact in forming the framework. This paper presents the components of a CIMF identified while working with the team of information managers within the LTER community.

2002 Proceedings of the 6th World Multiconference on Systematics, Cybernetics, and Informatics (July 14-18), Orlando, Fl.

2. ENVIRONMENT FOR A FRAMEWORK

System Drivers

The stimulus to share data (and ideas) in a broad context is, in many instances, conceived and implemented top-down by the higher administrative or intellectual hierarchy of an organization. In a local scenario, centralizing information in a Local Area Network (LAN) addresses the technical component of an organization's communication issues. When the need to become more efficient in the data management services provided to the community of users develops, a common framework provides a mechanism to scale from accessing and manipulating data to sharing information. In the case of the LTER, approaches to IM change as the research team's dedication to share data shifts from simply complying with sponsor directives to trying to integrate the different components of their research work. To act as a research team and to do synthesis, the scientific community members must be willing to share.

The LTER program sites adopt a holistic approach to local CIMF while addressing a broad scope of research activities [9] and sharing an articulated common IM vision [1]. The timeframe for data release depends on the complexity of the processes to gather, submit, and quality control the data. The LTER community responded to the sponsor's directive [15] to share data by preparing a data policy outlining data availability within two years of collection. In 1990 the LTER Information Managers Committee had reported several potential positive, negative and legal aspects of data sharing and recognized the value of making data accessible to others

---

Table 1. Data Access Policy

- *Identify existing community data set types according to their accesibility*
- *Agree on what data will be made available online*
- *Define data availability timeframe*
- *Define compliance guidelines for core and non-core data*
- *Provide justification for data sets not posted online*
- *Create written policy for secondary use of public data*
- *Document data use and data policy agreement*
- *Post a citation format, a disclaimer, and encouragement for ethical behavior to protect the data provider*

---

in order to advance science [14]. This document formed the initial guidelines for site information management policies. Table 1 summarizes some important aspects of data policy.

Infrastructure

Information management systems provide an infrastructure to serve the common interest of a research site's scientific community with tools to achieve synthesis and cross-site activities. When centralized, data generated in the field and laboratories along with its metadata are accessible to all participants (Strebel et al, 1994; Ingersoll et al, 1997; Brunt et al, 1998; Baker et al, 2000). Table 2 lists the infrastructure elements important in defining a complete CIMF.

Table 2. Infrastructure

- *A personnel directory to describe project participants*
- *A bibliography to list site related publications*
- *Metadata forms for data submissions*
- *A catalog of data sets consisting of data and metadata*
- *A description of projects for which datasets are generated*
- *In-house accessibility to data set owners while ensuring data integrity*
- *Data rights so owners have the option of exclusive access to their data*
- *Mechanisms giving file system access to local and off-site investigators*
- *Applications providing tools to extract, manipulate, and analyze data*
- *Back up system procedures for system security and disaster recovery*

Within this framework, the diversity of content and data formats must be captured by standardized data documentation (ISO, 2002; FGDC, 2002). The primary objective is that data be accompanied by metadata (data documentation) so that users are provided with the research context within which the data were created, the methodology used, and the specifications for each measurement [13]. Table 3 shows the relationship of the CIMF infrastructure elements, as defined and documented in the metadata, to the principal research components of the site. Each site adopts metadata standards to reflect its range of research activities. A set of minimum standards developed by the Network's Information Management Committee is available to all sites [16].

Community

The research community members can be divided into three main groups: a research team, an information management team, and an administration team. The administration team holds all the institutional administrative personnel who give support to the other two teams of the community. The research team (scientist, students, and field and lab technicians) has two distinct roles in a CIMF: the driving force for the ongoing revision of the infrastructure elements and the users whose needs are to be met. In an ideal situation the information manager works on infrastructure, computational, and dissemination tasks with the assistance of a staff of information specialists performing system and web design/ development/ maintenance, as well as data entry and programming. The information manager fills multiple roles within this scenario. Given the breadth of the responsibilities

Table 3. Infrastructure Organization

- *A cluster of two or more data sets could be defined by grouping related datasets within a broader research context*
- *When a cluster is defined, it can be described by answering the question of why the data sets are collected, why are they useful to the originator (the owner and creator of the data sets), and the scientific questions to which they are pertinent*
- *A title and the description of each cluster could be included in the metadata for each of the data sets involved. The description can point (link) to another metadata file containing the description*
- *The list of these clusters and the stand alone datasets provide a front end interface for the users visiting the data web page*

that must undertaken by the information management team, it is critical that the role of the information manager be well defined (Table 4).

Table 4. Role of Information Manager

- *Participate actively as member of local data manager committee overseeing information management priorities, policies, and compliance*
- *Participate in the design of data and metadata*
- *Agree upon the timeframe for data submission*
- *Establish the format or software with which data will be passed*
- *Identify who will do data processing, review, and quality assurance*
- *Define protocols for the role of field personnel in dataset creation*
- *Agree upon individual responsible for dataset document*
- *Prepare documentation standards*
- *Designate web page manager*

Although a CIMF coordinates a research team's data, the challenge of addressing human communication barriers still exists. A strong scientific background helps the information manager to better understand research needs and to maintain effective communications across a research group, but the role of the research team leader is critical in addressing the human communications dilemmas. Having a holistic view of the site's research work, the research leader can assist the information manager in identifying the critical components of the CIMF infrastructure that must be developed, enhanced or changed to better serve the continuously changing needs of the scientific community. This direct communication with the research leader of the community helps the information manager to better understand the vision and goals of the site's research. This interaction is critical specially when the information manager is not one of the community's researchers. In any situation, it is necessary for the information manager to have this vision for the design and implementation of a more efficient CIMF infrastructure. Also, the research leader could be the key element to resolve differences that might arise in regard to data management and sharing policy issues.

Another community challenge that the research and information management teams meet is to share data. The recognition of the need for a coherent information system as part of the organizational infrastructure increases with a better understanding of information flow. This understanding is blocked in an environment in which sharing jeopardizes a company's competitiveness in the market [10]. Resistance to sharing based on the potential for data misuse and misinterpretation is a common experience in other communities [17]. The fact that program sponsors require data sharing addresses this problem technically but not functionally. Given the contemporary consideration of intellectual property rights as well as the potential for data misuse and misinterpretation, data sharing requires not only attention from the information manager but the attention and insight of the research team leader as well.

3. TECHNICAL IMPLEMENTATION

Technical decisions impact data handling at all levels: entry, storage, conversion, processing, extraction, merging, and delivery. As an example, format is a central issue for data storage in an information system design. The decision to move data storage from a basic technical implementation to a more advanced approach adds complexity (raising significant support issues) while improving functionality. There are a variety of options for moving from flat ASCII-based files to

Table 5. Technical Elements

- *Establish where data archives will reside*
- *Create mechanisms to ensure security of the data*
- *Consider integration of personnel and bibliographic information*
- *Decide on storage format for metadata and for data*
- *Produce a data catalog that can be accessed from other information databases within the system*
- *Decide upon the role and design of the web page*
- *Create a virtual server address to give network identity*
- *Create mechanisms to facilitate the entry, update, and extraction of information such as data, metadata, publications and personnel*

relational databases today. The main purpose is to improve interface with the data and to provide an online dynamic rather than a static web interface (Table 5). Research groups often consider public domain, commercial software packages because of their low cost. Project choices are influenced by local computational infrastructure including: availability of platform support for Unix or NT systems, existing database licenses, existing file and web servers, and local expertise in interface languages like Lite, Perl or PHP. The choices involve an interplay of software elements including RDMS, web interface, and web server components (Table 6).

Table 6. Common RDMS/Web Implementation Options

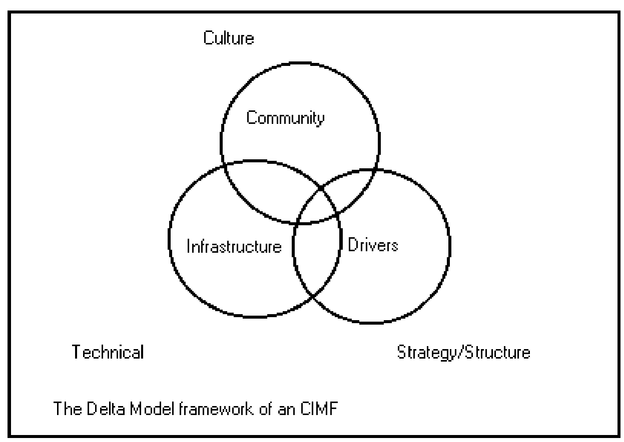
Platform	Web Server	Web Interface	DBMS
Unix	Apache	Lite	MiniSQL
Unix	Apache	Perl	MiniSQL, MySQL, Oracle
Unix	Apache	PHP	MiniSQL, MySQL, Oracle
Unix	Apache	Perl/DBI	Oracle, *others*
Unix	Apache/Tomcat	JSP (Java Servlet)	MiniSQL, MySQL, Oracle, Access, SQL Server
NT	Apache/Tomcat	JSP (Java Servlet)	MySQL, Oracle, Access, SQL Servert
NT	IIS	ASP	Access, SQL Server
NT	IIS	Perl/CGI	Access, SQL Server
NT	IIS	PHP	MySQL, Oracle
NT	IIS/Apache	Perl/CGI	Oracle

4. CONCLUSIONS

A common information management framework follows the Delta Model framework for Information Technology (IT) (Figure 1). This model has been used to induced organizational changes in which the components of technology (Infrastructure), strategy/structure (drivers), and culture (community) are considered critical for a holistic view when introducing IT to the community [2]. In a scientific community, a CIMF defines a research community and an infrastructure designed to facilitate the sharing and preservation of data for present and future generations (the goals of an LTER scientific community). A CIMF emerges in time as a result of the need of this community to make synthesis and cross-site studies. At the same time, the CIMF infrastructure promotes good science. By providing the needed technical (infrastructure) elements (centralized and integrated data set catalog, metadata, personnel data base, bibliography, research descriptions) it helps the research team and the rest of the scientific community to get a holistic view of their research and further promote synthesis and cross-sites studies. The teams' members (research, information management, and administration personnel) must aim to function in harmony with respect and mutual understanding. Even if this may be an idyllic situation, when the need to share information originates, people, especially scientists, work out their personal differences and do the work. The CIMF provides these teams with a set of common terms that facilitates technical discussions at the team level at the site, between the research and information management teams, and at the cross-site level, among the information managers.

Advanced techniques can provide more robust approaches to data query and retrieval for subsequent analysis and synthesis in a CIMF. These techniques can have a wide variety of technical implementations. Maintainability and sustainability of the CIMF software are other issues that must be addressed, in consideration of personnel and site infrastructure, as part of a total system assessment. A CIMF can be a evaluated by how well it serves its community in accomplishing its research goals. The research success of the LTER community, which manages an enormous amount of

Figure 1. The Delta Model framework of an CIM



site data in order to create relevant scientific discoveries [9], points to the LTER success at the information management level.

To be effective, a CIMF must be closely integrated with all components of an organization. It requires a commitment of resources and time not only from information managers but from the entire community.

5. ACKNOWLEDGEMENT

This work was supported by National Science Foundation (NSF) DEB-LTER 0080538, USDA Forest Service, and the University of Puerto Rico for the Luquillo LTER site, and NSF Grants OPP-96-32763, DBI-01-11544, and EIA-01-31958 for the Palmer site. We recognize the strength and uniqueness of the Long-Term Ecological Research program vision. We acknowledge the insights of the Luquillo and Palmer LTER site research communities and of the LTER information manager community past and present. Specific contributions by Donald Henshaw from the Andrews LTER site to form the common management framework concepts are acknowledged.

6. REFERENCES

[1] Baker, K.S., B.J.Benson, D.L.Henshaw, D.Blodgett, J.H.Porter, S.G.Stafford. 2000. Evolution of a Multisite Network Information System: The LTER Information Management Paradigm. BioScience 50: 963-978.

[2] Boekhoff, H. 1999. The Delta Model: A framework for the effective implementation of Information Technology to enable organizational change. World Multi Conference on Systematics Cybernetic and Informatics., Vol 1: 252 - 259

[3] Bowser, C.J. 1986. Historical data sets: lessons from the past, lessons for the future. In Pages 155-179, W.K.Michener , editor. Research data management in the ecological sciences. , University of South Carolina Press, Columbia, South Carolina, USA

[4] Brunt, J.W. 1998. The LTER network information system: a framework for ecological information management. In Pages 435-440, Aguirre-Bravo, Celedonio; Franco, Carlos Rodriguez, editors. North American science symposium: toward a unified framework for inventorying and monitoring forest ecosystem resources, 1998 November 2-6, Guadalajara, Mexico. Proceedings RMRS-P-12. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

[5] Federal Geographic Data Committee (FGDC). 1994. Content standards for digital geospatial metadata. Washington DC. http://www.fgdc.gov

[6] Ingersoll, R.C., T.R.Seastedt, M. Hartman, 1997. A model information management system for ecological research. BioScience 47: 310-316.

[7] International Organization for Standardization (ISO). 2002. http://www.iso.ch

[8] International Rice Research Institute (IRRI)'s Genetic Resources Center. 2001. Common Information Management System. http://www.irri.org/GRC/Datamanagement/IRGCIS.htm

[9] Kaiser, J. 2001, An experiment for all seasons. Science 293: 624-627.

[10] Lyyntinen, K. and D.Robey, 1999. Learning Failure in Information Systems Development, Information Systems Journal 9:85-101.

[11] Meléndez-Colom, E.C. 2001. A Definition of a Common Management Framework. http://luq.lternet.edu/publications/reports/informationmanagement/commonframe-version2.htm

[12] Michener, W. and J. Brunt, 2000. Ecological Data: Design, Management and Processing (Methods in Ecology). Blackwell Science.

[13] Michener, W., 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.

[14] Michener, W., Stafford S.G., and Nottrot, R., editors. Proprietary Issues. 1990. Pages 16 - 17 in Proceedings of the 1990 Data Management Workshop, Snowbird, Utah

[15] National Science Foundation (NSF). 2000. Long-Term Rcological Research (LTER): Special Competition for cross-site research. Program Announcement. NSF 00-61. http://www.nsf.gov/cgi-bin/getpub?nsf0061

[16] Porter, J. H., D.L. Henshaw, and S.G. Staford. 1997. Research Metadata in Long-Term Ecological Research (LTER). http://www.computer.org/proceedings/meta97/papers/jporter/jporter.html

[17] Rocheleau, B. 1997. Governmental information system problems and failures: A preliminary review. Northern Illinois University, Division of Public Administration, Public Administration and Management: An Interactive Journal. ISSN 087-0091. http://www.hbg.psu.edu/Faculty/jxr11/roche.html.

[18] Strebel, D., B.Meeson, A.Nelson, 1994. Scientific Information Systems: A Conceptual Framework. Pages 59-85 in W.Michener, J.Brunt, S.Stafford, editors. Environmental Information Management and Analysis: Ecosystem to Global Scales. Longdon: Taylor and Francis.

[19] Water and Sewer Board of Montgomery, Alabama. 2001. Total Utility Management System: a Common Information Systems. http://pasture.ecn.purdue.edu/~aggrass/esri95/