Home

Agenda

Participants

Resources:

Site Bytes

References

Presentations

Working Groups

Links:

LTER Home

LTER IM Page

IM2004 Report

IM2003 Report

Meeting Overview

The annual LTER Information Managers Meeting was held August 4-7, 2005 in Montreal, Canada and was attended by representatives from 25 of the 26 sites, members of the IT group from the LTER Network Office, and IT partners (link to participant list). The agenda included updates on current projects (items 2-6 below), working group discussions (7-13), and election of new IMExec members (14), as well as a video conference with Henry Gholz and Liz Blood at NSF. The first evening meeting had slide presentations by representatives from the two new LTER sites: California Coastal Ecosystem and Moorea Coral Reef as well as demonstrations of various software and websites under development. Special emphasis was given in the meeting to product-oriented working groups.

1. **Databits.** The current issue of the LTER Information Managers newsletter DataBits is available online at http://lternet.edu/databits. The newsletter is designed to engage the LTER IM community with a rotating editorship and authorship. Eda Melendez-Colon LUQ is editor for the fall issue. Brian Riordan BNZ is co-editor and will become editor for the spring 06 issue.

2. Cyberinfrastructure Planning. As part of the LTER Network planning grant, information managers from the network science working groups met with NISAC and the newly formed CI Team (Santa Fe, June 16-17, 2005) to initiate work on a strategic plan for LTER Network cyberinfrastructure (contact: Peter McCartney CAP).

3. ClimDB. Twenty-four of 26 sites now participate in ClimDB/HydroDB. Air temperature, precipitation and stream flow are the most commonly harvested variables. Sites are strongly encouraged to add other level-2 variables (especially relative humidity, solar radiation, and wind speed & direction) and data from early years wherever possible (contact: Don Henshaw AND).

4. IM Review Criteria. The IM review criteria approved at the spring 05 CC meeting were used in LTER site reviews this summer. An informal discussion and assessment at the IM meeting identified a few areas where the wording of the guidelines might be clarified, but on the whole the guidelines were found to be quite helpful both for reviewers and for sites preparing to be reviewed (contact: Emery Boose HFR).

5. Data Access Policy. Plans were discussed for implementing the

revised LTER data access policy approved at the spring 05 CC meeting. Sites will begin working to modify their data access portals to reflect the changes and accommodate the new standard criteria for data release (contact: Peter McCartney CAP).

6 . EML Implementation & Harvesting. More than 90% of LTER sites have implemented the EML standard for site metadata. About 75% of these sites have made their EML available to centralized servers such as the LTER MetaCat, and most of the remaining sites are close to doing so. There is still much variation among sites with regard to the percentage of datasets in EML, the level of EML implemented, and the percentage of EML harvested (contact: Inigo San Gil LNO).

7. SiteDB. This working group explored the use of the LTER site database (SiteDB; http://www.lternet.edu/sites/) to provide a single portal for uniform access to information, metadata, and data from individual sites. A series of improvements to SiteDB were proposed, including improved access, navigation, content, and links to individual site web pages. Sites are strongly encouraged to check and update the current information in SiteDB for their site (contact: Nicole Kaplan SGS).

8. Web Site Design. This working group continued a project initiated last year to develop recommendations for individual site web pages, especially for new sites and for sites redesigning their existing pages. The primary goals are to improve access to site information and to emphasize membership in the LTER Network. Once completed the recommendations will be circulated for comment, considered by NISAC, and presented to the CC for approval (contact: Nicole Kaplan SGS).

9. Unit Dictionary. This working group continued ongoing efforts to define common measurement units for data integration and synthesis. The Unit Registry Prototype (http://fire.lternet.edu/customUnit/) demonstrates how a dictionary of unit names, types, and definitions at multiple levels (site, working group, community, and domain) might be created by the community to facilitate generation of EML documents and help resolve syntactic and semantic ambiguities and conflicts (contact: Karen Baker PAL).

10. Controlled Vocabularies. This working group explored the use of hierarchical controlled vocabularies to facilitate the browsing and searching of LTER datasets. Keywords used to characterize most LTER datasets are currently uncontrolled and poorly suited for efficient searching. However, controlled vocabularies have been created for most scientific disciplines. The working group developed strategies for evaluating existing resources and possible collaborations to create a controlled vocabulary for the LTER Network (contact: John Porter VCR).

11. Attribute Ontologies. This workshop examined the potential benefits of creating attribute ontologies for LTER datasets. Ontologies are sets of related classes that provide semantic information for knowledge systems. Centrally-shared attribute ontologies could facilitate metadata entry and searching in the short run and data integration and synthesis in the long run. The working group outlined initial steps for creating a prototype to be evaluated in a workshop at the ASM meeting next year (contact: Peter McCartney CAP).

12. Community Standards . This working group discussed the process of creating community standards in information technology and lessons learned from the design and implementation of EML across the LTER Network (contact: Florence Millerand PAL).

13. All-Scientists Meeting 2006. This working group began planning for the ASM meeting in Estes Park next year. Recommendations will be developed for IM-related workshops as well as plenary talks, virtual poster sessions, and IT demonstrations. The IM meeting will likely be scheduled for 1-2 days before and/or after the ASM. Sites are encouraged to send their information managers to the ASM in light of the expected focus on cyber-infrastructure planning (contact: Jonathan Walsh BES).

14. IMExec Committee . Regular members for the coming year include: Emery Boose HFR, Corinna Gries CAP, Nicole Kaplan SGS, Eda Melendez-Colom LUQ, Ken Ramsey JRN, and Jonathan Walsh BES. Ex officio members include: Barbara Benson NTL, James Brunt LNO, and Don Henshaw AND.

2005 LTER Information Managers Meeting

August 4-7, 2005 in Montréal, Canada

"Enabling the LTER Decade of <Synthesis/>"

Home
Agenda
Participants

Resources:
Site Bytes
References
Presentations
Working Groups

Links:
LTER Home
LTER IM Page
IM2004 Report
IM2003 Report

Meeting Agenda

Date	Time	Meeting Topic (Speaker)
Aug 04 (Thu)	6:00PM - 7:00PM	Mixer and Dinner Buffet (Buffet starts at 6:30PM)
	7:00PM - 8:00PM	Introductions (Benson) Presentations (10 min each) Slide presentations by two new LTER sites: California Coastal Ecosystem (Jerry Wanetick, Lynn Yarmey) Moorea Coral Reef (Margaret O'Brien) Demo of Query interface to Metacat (Servilla) Demo of Unit Dictionary (Baker, Servilla) Demo IM Mentoring Site (White)
	8:00PM - 9:00PM	Open group discussion: 1) evaluation criteria in the context of this year's site reviews (Boose) 2) innovations at the sites
Aug 05 (Fri)	8:00AM - 8:15AM	Introductions and meeting goals (Benson)
	8:15AM - 8:20AM	Distributions of cyberinfrastructure needs assessment survey (Vande Castle)
	8:20AM - 8:50AM	NIS strategic plan (Brunt)
	8:50AM - 9:05AM	Status EML Implementation and Harvesting (San Gil)
	9:05AM - 9:30AM	Attribute Ontologies and Trends Project (McCartney)
	9:30AM - 9:45AM	Report: Website Design Groups (Kaplan)
	9:45AM - 10:00AM	Break

	10:00AM - 12:00AM	Working Group Session: Group 1: Web Access to Information: SiteDB (Kaplan, Baker, Melendez-Colom, Laundre, Gries) Group 2: Design of Query Interface to LTER Catalog: Controlled Vocabularies (Porter, Walsh) Group 3: Exploring Ontologies: Background (McCartney)
	12:00PM - 1:15PM	Lunch (IMExec meets)
	1:15PM - 3:15PM	Working Group Session: Group 4: Web Site Design Recommendations (Kaplan, Baker, Melendez-Colom, Laundre, Gries) Group 5: Unit Dictionary (Baker, Servilla, Yarmey, Powell, O'Brien, Sheldon, Haber, Millerand, San Gil) Group 6: Attribute Ontologies (McCartney)
	3:15PM - 3:30PM	Break
	3:30PM - 4:00PM	Teleconference with Henry Gholz and Liz Blood
	4:00PM - 4:15PM	LNO report on intersite databases (Brunt)
	4:15PM - 5:00PM	LTER Network Planning Grant: Cyberinfrastucture (Benson, McCartney)
	5:00PM - 5:30PM	Break (cash bar)
	5:30PM - 6:30PM	Working Group Session: Group 7: Community Process and Standards Implementation (Baker, Benson, Jones, Millerand)
Aug 06 (Sat)	8:00AM - 9:45AM	Plenary: Working Group discussion (Powell, moderator)
	9:45AM - 10:00AM	Break
	10:00AM - 12:00PM	Working Group Session: Group 8: Joint: Ontology, Unit Dictionary, Controlled Vocabulary Groups (Baker, Servilla, McCartney, Porter, Powell, Yarmey) Group 9: All Scientist Meeting 2006 Planning (Ramsey, Kaplan)
	12:00PM - 1:15PM	Lunch (IMExec meets)

	1:15PM - 2:15PM	Business Meeting (closed) Election of IMExec members Databits report and guidelines Discussion of method to select NISAC members
	2:15PM - 2:45PM	NEON presentation (Michener)
	2:45PM - 3:00PM	Break
	3:00PM - 3:30PM	Discussion: implementation of the new data access/data use policy
	3:30PM - 5:00PM	Reports and Discussion (presentations each no more than 15 minutes to leave time for discussion) LTER Grid Prototype (Servilla) Sensor Network Grant: Automation, Scaling (Benson) SEEK (Jones)
Aug 07 (Sun)	8:00AM - 10:00AM	Plenary: Working Group reports and discussion (Sheldon, moderator)
	10:00AM - 10:15AM	Break
	10:15AM - 11:00AM	Discussion of NIS strategic plan and long range planning (Henshaw)
	11:00AM - 11:30AM	Identification of action items (Benson)
	11:30AM - 12:00AM	Meeting wrap-up (Benson) Meeting report

Meeting Participants

LTER Sites

AND - Don Henshaw, Suzanne Remillard, Theresa Valentine
ARC - Jim Laundre
BES - Jonathan Walsh
BNZ - Brian Riordan
CAP - Peter McCartney, Corinna Gries
CCE - Jerry Wanetick, Lynn Yarmey
CDR - Steven Bauer
CWT - Barrie Collins
FCE - Linda Powell
GCE - Wade Sheldon
HBR - John Campbell
HFR - Emery Boose
JRN - Ken Ramsey
KBS - Sven Bohm
KNZ - Jincheng Gao
LUQ - Eda C. Melendez-Colom
NTL - Barbara Bension, David Balsiger
NWT - Todd Ackerman
PAL - Karen Baker, Florence Millerand, Shaun Haber
PIE - Robert Garritt
SBC - Margaret O'Brien
SEV - Jim Gosz, Kristin Vanderbilt
SGS - Bob Flynn, Susan Stafford
VCR - John Porter

LTER Network Office

James Brunt
Jeanine McGann
Inigo San Gil
Mark Servilla
Duane Costa
John Vande Castle
William K. Michener
Marshall White
Bob Waide

Invited Guests

Avinash Chuntharpursat (CERN)
Judy Cushing (Canopy Database Project)
Matt Jones (NCEAS/SEEK)
Judith Kruger (Kruger National Park)
Sheng-Shan Lu (TERN)
Mark Schildhauer (NCEAS/SEEK)
Burl Timothy (Tim) Rhyne (NBII)
Ferdinando Villa (UVM/SEEK)

Home

Agenda

Participants

Resources:

Site Bytes

References

Presentations

Working Groups

Links:

LTER Home

LTER IM Page

IM2004 Report

IM2003 Report

LTER Site Bytes

- AND (Andrews LTER)
- ARC (Arctic LTER)
- BES (Baltimore Ecosystem Study)
- BNZ (Bonanza Creek LTER)
- CCE (California Current Ecosystem)
- CAP (Central Arizona - Phoenix)
- CWT (Coweeta LTER)
- FCE (Florida Coastal Everglades)
- GCE (Georgia Coastal Ecosystems)
- HFR (Harvard Forest)
- HBR (Hubbard Brook LTER)
- JRN (Jornada Basin)
- KBS (Kellogg Biological Station)
- KNZ (Konza Prairie LTER)
- LUQ (Luquillo LTER)
- MCM (McMurdo Dry Valleys)
- MCR (Moorea Coral Reef)
- NWT (Niwot Ridge LTER)
- NTL (North Temperate Lakes)
- PAL (Palmer Station)
- PIE (Plum Island Ecosystem)
- SBC (Santa Barbara Coastal)
- SEV (Sevilleta LTER)
- SGS (Shortgrass Steppe)
- VCR (Virginia Coast Reserve)

LTER Site: Andrews LTER

Contributor: Suzanne Remillard and Don Henshaw (Aug 02, 2005)

Site Byte:

Information Management (IM) folks at the Andrews Experimental Forest have spent the past year preparing for the mid-term review. We are pleased to announce that we had a very successful and positive review in mid-July. Our Information Management System was emphasized as one of the exemplary pieces to the Andrews LTER. Descriptions of our system are online at http://www.fsl.orst.edu/lter/research/component/infomgt.cfm?topnav=63

The IM team is led by Don Henshaw who has been heavily involved with cross-site synthesis and network-level efforts, and has also guided site preparation for the mid-term review. Gody Spycher continues to populate our SQLServer metadata database with remaining core and other legacy databases, develop programs for QA and other maintenance functions, and work with scientists to cleanup and analyze data. Gody plans on retiring within the next year, and we will begin to strategize about IM at the Andrews without Gody. Suzanne Remillard is recently hired as the Andrews LTER Information Manager and will begin taking over duties from Gody. Suzanne continues as the database administrator for the combined ClimDB/HydroDB effort, and actively maintains our web pages and attends to many data processing and IM tasks. Theresa Valentine continues to work with spatial data and is incorporating the HJ Andrews LTER spatial data into our general data catalog Theresa has also received seed money from Forest Service R&D to continue development of WatershedDB, a collection of spatial data for HydroDB sites including research area and watershed boundaries, gaging station locations, stream networks, and DEMs. Theresa also published a new, updated Andrews Experimental Forest map (http://www.fsl.orst.edu/lter/about/site/map.cfm?topnav=157).

Andrews Databases Online / EML: We currently have 130 databases (metadata and data) online and available. This includes over 500 data tables. Of these databases, 18 are spatial databases, with 50 more in preparation (300 coverages / Shape files). Our maintenance tasks are heavy, with over 100 tables updated annually. In addition, we receive about 10 new databases per year. We have only 6 databases left to migrate to SQLServer and an additional 40 legacy databases not online (to be migrated too). Our metadata database dynamically generates EML, and also produces downloadable PDF files for each study database. We are now contributing EML to the LTER Metacat through dynamic harvest list and EML generation scripts. Our next major focus is to review our EML with respect to accepted syntax and best practices, and revise our EML generation accordingly.

Reservation program: Gody and Suzanne have developed a new reservation system for the Andrews field station. This system not only enables the tracking of users, but also allows for the management of billing. There is also an online reservation request form.

Andrews goes wireless and adds spread spectrum telemetry
Fred Bierlmaier, the on-site Andrews system administrator, maintains the site Local Area Network (LAN), local web server, wireless LAN, digital radio and spread spectrum telemetry networks, telephone communications, and local personal computers. Fred recently installed the wireless LAN which links the dormitories, cafeteria, shop, and director's residence to the wired LAN with a wireless bridge. Wireless access points are also installed in the conference hall and throughout most

of the compound, and access points are planned for the classroom and library. A telemetry system based on 900 megahertz spread spectrum radios was installed in 2004 to access specific study data coming from data loggers on a nearby watershed.
LTER Graduate Student Collaborative Symposium: The first LTER Graduate Student Collaborative Research Symposium was held on April 13th through the 17th at H. J. Andrews LTER. This symposium was created in order to facilitate future graduate student interaction and participation in the broader community of LTER scientists, as well as to stimulate graduate student engagement in comparative and collaborative research efforts. There were a total of 66 graduate students representing 24 United States LTER sites and 11 students from international LTER sites (representing China, Mongolia, South Africa, Austria, Czech Republic, Brazil, Mexico and Switzerland). There were also 2 invited speakers (Dr. Whendee Silver and Dr. Scott Collins), 2 post-docs, 1 undergraduate, and multiple LTER information managers in attendance. Overall, there were 71 presentations (including LTER site review presentations and personal research presentations) and 13 training and collaborative workshops. The workshops in addition to being very informative also sparked at least five long-term cross-site collaborations. These collaborations are using long-term datasets that are available from LTER sites and will be used for cross-site comparison and eventually appearing in peer-reviewed publications. The most rewarding aspect of the event at its close was the feeling that we are not simply working at individual LTER sites but that we are all part of a large and warm community—an LTER network.

LTER Site: Arctic LTER

Contributor: Jim Laundre (Aug 03, 2005)

Site Byte:

The Arctic LTER now has all of it legacy metadata converted to EML. With the help of Duane Costa, Mark Servilla and Inigo San Gil we have successfully converted the structured text files to level ~ 2.5 eml and harvested the files to metacat. One more revision will be done to add a complete title to the files. Our original metadata files did not have titles and the current eml files use the file names as titles. Moving forward we are developing an Excel entry sheet for entering metadata information. The metadata sheet can be copied into to the researcher’s excel workbooks and included with their data. An Excel VBA macro will be used to output an eml file.

The current field season is going well. Spring was early with ice cover on Toolik completely gone by June11th. July has been cool with some occasional snow. This year there will be power and internet connection during the winter months at Toolik Station. With year-round power we

have made the current weather data available online, which can be viewed at http://ecosystems.mbl.edu/arc/ .

LTER Site: Baltimore Ecosystem Study

Contributor: Jonathan M. Walsh (Aug 03, 2005)

Site Byte:

BES Site Byte 2005-08-03

Jonathan M. Walsh

- We are approaching 5 years of online stream chemistry, and 4 years of online bird survey data
- We have EML level one capability. We have not reached level one compliance just yet because we wish to devise a means to ensure revision level control for our metadata before we expose it to the internet.
- We have a geodatabase running. It utilizes arcsde and db2 running on Linux.
- Metadata stored on our rdbms system will automatically populate our static "Data for download" page.
- We have begun designing experiments relating to urban design and the ecosystem and have formed an Urban Design Working Group.
- The third Vital Signs publication has been released. Brought to you by the Baltimore Neighborhood Indicators Alliance, the publication takes the pulse of Baltimore utilizing 40 indicators. The indicators fall into seven main categories ranging from housing, sanitation, and safety to workforce, health, environment, and child well-being. The Vital Signs are also featured in the Baltimore Sun every Monday. (See www.bnia.org)
- Watershed 263 Survey: We are surveying all BES-Affiliated researchers to gain information that describes the breadth of our research connected to Watershed 263. (Watershed 263 is a watershed comprised wholly by man-made waterbodies and has no "natural" water body) An online survey was developed for this purpose.
- Data management survey: A survey to study data management activities among BES researchers was completed. Results are forthcoming.
- BES Publications online: The BES publications have been imported into a database and are now searchable online. See http://www.beslter.org . There is a button on the top to access the publications database.

LTER Site: Bonanza Creek LTER

Contributor: Brian Riordan (Aug 01, 2005)

Site Byte:

It has been a busy half a year at BNZ. I took over the reins as Data Manager on Dec.1st. Since then I have redesigned our database for a flat file mysql database to a relation database supporting views, triggers, and stored queries. In addition, we rebuild the servers to have up to date hardware and software. I have taken all of the old ascii dataset and generated mysql tables for each one. This will allow us to generate complex searching capabilities as well as ensure better data-backups. We have installed a robust back-up unit that now backs up our servers as well has the majority of the BNZ staff.

We are in the process of generating full PDF's of our publications as well as deploy a more error free searching format for locating publications.

The next step is to redo our Coldfusion queries to generate results form the new database.
In addition, we are exploring the best way to generate our EML out of this new format. Currently we have little to no metadata in EML format. I have handed out the Excel worksheet that FCE created and the PI's have responded well to it.

All in all it has been a year of learning for me. I have gained a lot of knowledge about the BNZ database/website as well has how to make it better. The next year will be focused on deploying the database and generating EML for each data file. If this EML will be generated dynamically or stored as hard files is still to be decided.

See everyone soon!

LTER Site: California Current Ecosystem

Contributor: Karen Baker (Aug 27, 2005)

Site Byte:

In its first year as an LTER site, the CCE information management effort focused on building technical, organizational, and social infrastructure. The use of mail, storage and server facilities are coordinated centrally within the Integrative Oceanography Department at SIO/UCSD. Participation in an Ocean Informatics Environment is providing a contemporary approach to training as one dimension of design team and working group activities.

Initial projects included design of a web site (http://ccelter.sio.ucsd.edu) and establishment of the LTER network virtual pointer (http://cce.lternet.edu). A new shipboard organization scheme has been deployed, centered around event numbers as a long-term integrative element. Work to establish dictionaries and controlled vocabularies is proceeding in parallel

with design of metadata forms. From discussions, an emerging common language includes critical informatics terms such as data types, core data, metadata, and integrative indexes. A site software and data survey is being conducted as a part of the process of establishing a culture of LTER data practices. This represents an opportunity for learning about existing data handling within local environments.

Collaborative local activities included coordination with the Palmer LTER, the Southern California Coastal Ocean Observing System (SCCOOS), the Palmer LTER and the California Cooperative Oceanographic Fisheries Investigations (CalCOFI) program which provides an ongoing 50 year field time series for a grid encompassing the CCE sampling lines as well as with the Palmer LTER. As LTER Network participants, we contributed to several working groups at the 2005 annual IM meeting: Dictionary Process Unit Repository; SiteDB and Web Design; and Community Process. In collaboration with science studies participants, we are working to open up discursive data practices and technical perspectives by incorporating collaborative design as one element of focus for a comparative environmental research study. We are using this approach as a powerful tool to address immediate needs and choices within information management in general and standards' development in particular.

LTER Site: Central Arizona - Phoenix

Contributor: Corinna Gries and Peter McCartney (Aug 08, 2005)

Site Byte:

This year at CAP the main web site saw some improvements and additions. Overall programming changes were implemented to increase the speed at which database searches retrieve and display results. CAP's new 'Integrated Project Areas' are described and associated subprojects and people are displayed in the same way as LTER core areas. An image library with search interface was added and currently we are working on a CAP sites atlas. An application built on ArcIMS/html viewer that allows users to display a map with the locations of all sites at which research activities are taking place within the CAP area. This will enable the user to locate sites at which for instance all long term monitoring programs are currently active. With appropriate background layers other questions may be answered, e.g. in which watershed a site is located or which soil conditions are prevalent at a certain site, etc. This application was developed in collaboration with the ASU GIS lab.

However, the main thrust this year was finishing the programming for an intranet (http://iis.asu.edu/intranet). This web application allows CAP researchers in a secure environment to enter and edit their personal information, their publications, calendar events, information on datasets

they are producing, and information for each subproject. For each subproject beginning and end, a short description, associated people, and an annual report can be entered. Although we are still working out bugs this application has been successfully employed to gather information for this year's annual report to NSF. With this year's experiences further improvements, especially in the area of editing and bulk uploading of publications will be implemented next year.

The Center for Environmental Studies, where CAP LTER is housed, became the International Institute for Sustainability with a large private endowment to help promote ASU's development of a strong research and educational mission in sustainability.

ASU was awarded a 5 year grant for the "Decision Center for a Desert City". This project compliments CAP LTER by providing a strong outreach to the decision-making community in central Arizona. Funded by the NSF Decision Making Under Uncertainty program, the project focuses on the social and ecological issues surrounding long -term water management in the Phoenix basin. Plans are under development this summer for a tri-university partnership in a distributed water information system which will expand the core data archive that was started by CAP LTER to a state-wide integrated system.

IIS completed its NSF ITR project "Integrating Urban Models Through Network Services". This project focused on developing solutions for integrating distributed models maintained by different research and management institutions within the valley. The project developed web service wrappers to Grass and R software packages, web interfaces for model execution requests, extensions to EML metadata for documenting model inputs and outputs, and a set of actors for Kepler designed to interface with our Xylopia web services. Development started under this project will continue as part of the DCDC project.

IIS Informatics lab collaborated with our Departments of Anthropology and Computing Science on a successful NSF Human and Social Dynamics proposal on developing a national cyberinfrastructure for Archaeology. This pilot project supported a workshop at NCEAS in December which had participation from SEEK and GEON personnel, plus a small pilot study using OWL ontologies to guide automated reclassification of archaeological datasets. A larger proposal is now being considered by the NSF CyberTools program which, if funded will allow IIS to expand and refine the work we have done with EML, Xanthoria, etc to a broader disciplinary base.

Robin Schroeder (whom many of you know as Robin Schoeninger) has taken a position as assistant curator for the ASU lichen herbarium and data manager for all ASU natural history collections. She will continue to collaborate with our lab helping to maintain and expand our internet biodiversity applications. An advertisement for a new programmer

position at IIS will be forthcoming.

The ASU GIS lab has moved to a new home with the ASU Institute for Geography. Again, we are expecting the fruitful collaboration to continue.

LTER Site: Coweeta LTER

Contributor: Barrie Collins (Jun 20, 2005)

Site Byte:

Coweeta LTER IM will be part of our mid-term review June 28-29, 2005. The past year has been a continuance of our infrastructure development. Key accomplishments include bringing online 1100 publications (full pdf form), internet map services, GIS for the entire southern Appalachian study region, as well as full participation in EML and the Metacat Data Harvester.

Our focus in the coming year(s) will be to continue to link data, publications, research, researchers, and spatial data to increase the stream of information. Navigational structures have been brought into the 21st century as we endeavor to understand how to serve a very deep web site to a varied user base efficiently and effectively.

That's about all for now, I reckon.

Barrie

LTER Site: Florida Coastal Everglades

Contributor: Linda Powell (Aug 03, 2005)

Site Byte:

Over the past year, Linda Powell (information manager) and Mike Rugge (project manager), have concentrated their efforts on several important components of the Florida Coastal Everglades (FCE) LTER program's information management system: 1) designing and implementing a new tool to capture EML, 2) expanding the Oracle9i Database and 3) upgrading the FCE web page.

One our site's of the most significant accomplishments has been the creation of a new EML tool called 'Excel2EML'. Since the inception of the FCE LTER, comprehensive metadata for each dataset has been entered into an Excel metadata template by our individual researchers. The challenge for the FCE was to find a way to capture EML compliant (tier 4 or 5) metadata using the Excel spreadsheet format our researchers were happy using and to convert it to a valid EML XML document. Mike

Rugge spent several intensive months developing and tweaking the Excel2EML tool and in November of this past year, the tool was made available to the IM community through a LTER network office (LNO) CVS (http://cvs.lternet.edu/cgi bin/viewcvs.cgi/eml/tools/Excel2EML/). We were able to add a MAC OS X version of the tool to the CVS in February 2005. Links to the Excel2EML CVS can also be found in the Data section of the FCE website at http://fcelter.fiu.edu/data/tools/. Three quarters of the FCE legacy metadata have been re-entered into the new EML metadata templates and converted and will be ready to be harvested by the LNO metacat by the third week of August 2005.

In the upcoming weeks, we will be upgrading our information management system by adding three large servers to replace existing equipment: 1) a web server, 2) an Oracle9i database server and 3) a data server. Our plan is to integrate our project management information, data and corresponding metadata and project GIS coverages into our Oracle9i database. We still have a few web pages driven by our original MS Access database and Mike Rugge is nearly finished migrating this project information into the Oracle9i database. We have been busy collecting and entering LTER and LTER related project information into our Oracle9i database in order to enhance our web-based interactive mapping application called the 'FCE LTER Interactive Everglades Map' found on our web site at http://fcelter.fiu.edu/gis/everglades-map/.

As we are beginning to prepare for our 2006 funding renewal, several enhancements have been made to the FCE website. We've slowly been adding detailed information into our 'researcher profile' pages and have expanded personnel queries to include FCE site roles, research interests, FCE working group membership, affiliated organization, and keywords. Mike Rugge has improved the FCE publications web page, allowing the user to query our archive by author, journal name, publication type, and keyword. In the past year, we worked with Susan Dailey (FCE Education and Outreach Coordinator) to add a FCE Schoolyard program component (http://fcelter.fiu.edu/schoolyard/) to the website. We plan to work with the FCE Graduate Student Organization to update their section of the website.

LTER Site: Georgia Coastal Ecosystems

Contributor: Wade Sheldon (Jul 31, 2005)

Site Byte:

We are already busy preparing for our 2006 renewal proposal, so information management work at GCE has largely been focused on helping PIs with data synthesis projects, improving managment and access to GCE reprints and publications, and tightening integration

between project databases through web application cross-references and content displays. The data synthesis efforts have been strongly aided by the development of the GCE Data Search Engine, a MATLAB application for metadata-based indexing, searching and integration of data sets stored as GCE data structures (http://gce-lter.marsci.uga.edu/lter/research/tools/toolbox_search.htm). This tool has been well recieved by PIs, and several use it to manage and analyze their own data in addition to GCE data sets. We also began using this indexing technology to generate detailed data summaries to enhance distribution of near-real-time and historic ancillary data on the GCE Data Portal web site (http://gce-lter.marsci.uga.edu/portal/). This is particularly useful beause we began providing full public access to the portal site this year to support our SLTER participants and UGA students as well as the broader LTER community. All portal data sets can be searched and retrieved alongside public data from the GCE Data Catalog using the GCE Data Search Engine client (i.e. via direct HTTP in MATLAB). All data access from the GCE Data Catalog and GCE Data Portal and all software downloads are now tracked in our data access database, requiring contact and affiliation information from downloaders, but we did add support for web browser cookies and stored logins in the Search Engine client to streamline the downloading process for regular data users.

We are holding off on implementing new data search applications on the GCE web site until the standardized LTER query interfaces in development are more mature, but we did significantly enhance the query capabilities of our existing data catalog interface this year (http://gce-lter.marsci.uga.edu/lter/asp/db/data_catalog.asp). Specifically, we added support for searching by LTER core area, searching by study dates, and also by text in individual or combined metadata sections. Metadata content searches are highly targeted and extremely fast compared to xml-based searches, and the searchable content can easily be tailored in the future by creating simple database views and adding them to the array of available metadata sections. These new query capabilities also allow search hyperlinks to be added anywhere on our web site to support improved cross-referencing between databases, such as searching for relevant data sets for individual species records in our taxonomic database (e.g. http://gce-lter.marsci.uga.edu/lter/asp/db/species_details.asp?id=Geukensia%20demissa)

This year we began acquiring and managing more full text reprints for GCE publications and we now have a searchable reprint archive linked to our bibliographic database on the private GCE web site for project participants. We also automatically generate downloadable reprint archives complete with HTML indices of all reprints (both with and without abstracts) to allow PIs and students to install up-to-date libraries of GCE publications on their local systems for research purposes. We also improved linkages between our bibliographic and personnel databases and began providing an automated reprint request service on

the public web site for both GCE and UGA Marine Institute publications (http://gce-lter.marsci.uga.edu/lter/asp/db/biblio_query.asp). Reprint requests are emailed to first listed GCE-affiliated author and cc'd to the IM office, and if no authors are active GCE participants the request is directed to the IM or UGAMI librarian instead. Quite a few international researchers and students have been using this service to request reprints from journals that are not widely available overseas.

We continue to provide complete Level 5 EML metadata for all data sets added to the GCE Data Catalog, as well as customizable versions of species lists in EML format from the GCE taxonomic database (http://gce-lter.marsci.uga.edu/lter/asp/db/all_species_lists.asp). We improved our EML implementation to better match the current Best Practices recommendations this year (e.g. reorganized geographic coverage content to describe the overall bounding box under eml/dataset/coverage, with individual site descriptions moved to the eml/dataset/methods/sampling/studyExtent tree, as well as inclusion of standard "place" keywords). All dynamically-generated EML documents are harvested weekly for inclusion in the LNO and KNB metacats and also the NBII metadata clearinghouse.

On the IT front we experienced our first major hardware problem this year when a SCSI hard drive in our database server's RAID-5 array failed and an auxiliary drive developed some bad sectors. The RAID did its job and kept the system up until new drives could be obtained, but that process turned out to be very complicated (and expensive) due to the age of the drives, taking over six weeks. After this experience we decided it was time to acquire a new database server, so I was able to purchase a PowerEdge 2800 with dual Xeon cpus and a faster RAID-5 array and re-task the older server as a development box. We also tightened security for accessing all servers by generating SSL certificates and requiring HTTPS (or SSH/SFTP) for all authentication. Less secure transport protocols (e.g. FTP, WebDav) are now disabled on all our systems and all machines are protected by firewalls (in addition to the rule-based URL filter on the web server). This required some PI hand-holding, particularly installing and using SFTP software, but should help us minimize our "attack surface".

We also revised our project management structure this year by forming an executive committee to direct the project, and it was unanimously decided that the information manager should serve as a full voting member on this committee. This formal acknowledgement of the central role the IM program plays at our site is very encouraging, and should improve the already strong integration of our science and IM programs in long term planning. In addition to forming an executive committee we also established formal bylaws for GCE, and we developed a secure web application for voting on project referenda in accordance with these bylaws. This application allows our broadly distributed PIs to vote on important issues that arise between annual meetings and will also help us

establish a formal record of major project decisions for posterity.

LTER Site: Harvard Forest

Contributor: Emery Boose and Julie Pallant (Sep 15, 2005)

Site Byte:

The functionality of our web page was greatly increased this year through incorporation of online databases in MySQL. This addition has streamlined our application process for on-site research, fellowship positions, and undergraduate (REU) students. In addition, we are posting more dynamic web pages with information submitted to our databases. The 16th annual Harvard Forest LTER Symposium abstracts were submitted and posted entirely on-line.

There is growing interest at our site in extending wireless networking to our main research area (Prospect Hill Tract, 1000 acres). Some interesting technical challenges are presented by the local topography and dense forest (not to mention hurricanes, ice storms, etc). We are also designing and installing a series of long-term hydrological stations across Prospect Hill. One of the goals of the wireless network would be to permit near-real time collection, processing, and display of hydrological data, as we currently do for meteorological data (http://harvardforest.fas.harvard.edu/hfmet/).

Miscellaneous items - Major renovations to our main building (Shaler Hall) involved installation of new network and telephone wiring as well as a new audio-visual system. Great strides were made in the reorganization and classification of the Harvard Forest Library and inclusion of our holdings in the University's on-line catalog, with generous help from professional librarians at Harvard. A major lightning storm in July caused extensive damage to our phone system, meteorological station, eddy flux towers, and soil warming experiment (whose instrument shack burned to the ground).

We are currently updating our information management system in preparation for our LTER renewal submission in Feb 06. Updates will include data and metadata from our Schoolyard LTER program last year. Plans for the beginning of LTER IV (assuming we're funded) include migrating our scientific metadata from EML to MySQL, migrating core datasets from flat files to MySQL, and incorporating more spatial data into our information management system.

LTER Site: Hubbard Brook LTER

Contributor: John Campbell (Jul 29, 2005)

Site Byte:

We have made significant progress with our EML implementation. We now have 115 EML documents harvested into the metacat. All of our non-geospatial datasets were converted to EML last year, and this year we converted all of our spatial metatadata to EML. We used the Esri2Eml stylesheet that was developed at CAP and are grateful for help from Corinna Gries. A stylesheet was developed for displaying metadata for non-geospatial data and we will incorporate this into our web page shortly (see http://www.hubbardbrook.org/eml/generate_html_from_eml.php). We also plan to develop a separate stylesheet for displaying metadata for spatial datasets.

We are in the process of completely redesigning our web page and have made considerable progress on this front. A major part of this effort involves redesigning the way data are queried and downloaded. We have moved all our data into a MySQL database and will query it with php applications (see example at: http://www.hubbardbrook.org/search_portal_db.php). We also linked the EML based metadata to the new LTER network data policy and will register users to track data downloads see (http://www.hubbardbrook.org/data_policy.php?target=801d8c3700665679e56756c2b8ab84081b).

During the last year we imported our site bibliography into Endnote to remain consistent with the LTER network office. The bibliography contains nearly 2000 publications that have been published since the inception of the Hubbard Brook Ecosystem Study. We will continue to develop the bibliography and plan to modify the query interface on our web page.

In the last year, there has been considerable interest in implementing wireless technology and real-time data at Hubbard Brook. Past efforts have been hampered by a number of complications including mountainous terrain, remote location, dense forest cover, and a lack of electricity. We are currently exploring practical and cost efficient ways to overcome these challenges and are in the proposal writing phase.

The Hubbard Brook information management team consists of Ellen Denny, Netta Kies, Phyllis Likens and John Campbell. Ellen works at the USDA Forest Service in Durham and helps maintain the Hubbard Brook web site and has been working on the redesign. Netta Kies is a summer employee at the Forest Service and will return to Wesleyan University for her senior year in the fall. Netta has been developing the web database and associated php applications. Phyllis Likens is an employee of the Institute of Ecosystem Studies in Millbrook, NY. Phyllis compiles the Hubbard Brook publication list, maintains the document archives, and helps with the personnel database and curriculum vitas. John Campbell continues in his role as Information Manager for the Hubbard Brook

LTER working on a variety of tasks

LTER Site: Jornada Basin

Contributor: Ken Ramsey (Aug 02, 2005)

Site Byte:

The information management team of the Jornada Basin LTER has been very busy in the last year responding to research needs, creating Geographic Information System (GIS) products, implementing Ecological Metadata Language (EML), application development and deployment, database population, effectively doubling server storage capacity, and participating in LTER Network planning activities. The information management team consists of Barbara Nolen (GIS/Remote Sensing), John Anderson (Site Manager), Ken Ramsey (Information Manager), and Justin Jensen (student programmer). The information management team receives network and desktop systems support from Jim Lenz of the USDA ARS Jornada Experimental Range (JER).

Barbara Nolen has had a productive year. Her products this year includes a broad scale map showing major JRN and JER research sites, a Schmidt's delineation of the Chihuahuan Desert, digital elevation models of the Chihuahuan Desert and New Mexico, and LandSat images of the Chihuahuan Desert and New Mexico. Barbara is currently working on a tour book for the World Congress of Soil Science and mapping landforms of the Chihuahuan Desert and New Mexico.

The Jornada Basin LTER hired a student programmer, Justin Jensen, this year to develop PHP and Java applications for the Jornada Information Management System (JIMS) databases. With support from Central Arizona Phoenix LTER (CAP) and the LTER Network Office (LNO), Ken and Justin went to Tempe, AZ for assistance in implementing Xanthoria to generate EML from the JIMS database tables. Corinna Gries (CAP), Robin Schroeder (CAP), Ken, and Justin spent 3 days in Tempe developing the initial implementation of Xanthoria that generated most of the major elements of level 5 EML based on LTER Best Practices. Justin and Ken completed the XSL style sheets and Xanthoria configuration file following the trip to CAP.

Ken is currently completing the population of the JIMS databases with data, associated metadata, and personnel information needed to generate EML and develop and implement a new content management system for a new dynamically generated web site currently under development. This includes associating people with research projects and datasets within related database tables. JRN is set to deploy a searchable people directory, dynamic data catalog, and semi-automated LTER ClimDB harvest mechanism (68+ stations) following internal revue of the

interfaces. As the JIMS databases are populated with a research project, EML instance documents will be generated, archived locally, harvested to the LNO Metacat node, and made available on the JRN web site.

JRN has developed a set of rules for generating EML based on the JRN IMS database schema and LTER EML Best Practices. John's attendance of an EML workshop at LNO prior to this planning process was very helpful in allowing us to quickly determine JRN rules for EML content. The LTER Best Practices Document was crucial in guiding JRN during the process of developing and implementing EML.

Jim and Ken have made several improvements to the IMS servers and services in the last year. The total storage capacity for IMS servers was doubled by adding a 2 TB RAID enclosure and reallocating storage capacity among the file, database, and web servers. This brings the combined server capacity to just over 4 TB. The web server was moved to a newer server with SUSE Linux as the operating system and Apache web server. The operating system for the file server was upgraded to Netware 6.5 during the storage capacity reallocation. The Internet map server has been upgraded and is now performing at a significantly increased level of performance.

John has recently added 5 mini-flume hydrological research sites to the Jornada wireless system. John has installed the new LTER weather station and is currently finalizing its configuration and calibrations. The new weather station has also been added to the Jornada wireless system.

LTER Site: Kellogg Biological Station

Contributor: Sven Bohm (Aug 03, 2005)

Site Byte:

With the EML delivery well under way, we concentrated this last year on cleaning up and refactoring the data delivery code, and work towards removing the jsp pages in favor of testable java modules. We also started to develop some tools to help users do quality control, and some data processing for the longer term data. This will hopefully allow us to capture data earlier, when memories of the metadata are still fresh.

The goals for this year are to bring our air-photos online and integrate the spatial data points (field corners, sampling stations and the like) with the traditional data. Our new project server is finally online, with the old server used as a "live backup". We have started linking the publication database to the datasets.

Thats about it.

LTER Site: Konza Prairie LTER

Contributor: Jincheng Gao (Aug 03, 2005)

Site Byte:

There have been lots of changes at the Konza site in the past year. We rebuilt our Local Area Network (LAN), updated our SQL Server and web site. In spring 2005, I took over the position of KNZ Information Manager and continued to work on data quality control, database population, and interactive maps of spatial data to prepare for the mid-term site review in early June. I am glad to let you know that we had a successful and positive review, and also received some good and constructive suggestions for the continuing development of our IM system.

A wireless network was installed at the Konza Prairie Biological Station last year that facilitates transferring data directly from the station to a data server on campus. A local network system was rebuilt and the network operating system was transferred from Novell to Window 2003. Symantec antivirus server was installed for retrieving and distributing virus signature and program updates. A new database server, SQL Server 2000, was established last year. All KNZ LTER metadata and data are stored and managed on the server. Right now, most metadata are stored in three formats- text, PDF, and EML. Most datasets are stored in both text files and in the SQL Server database. Metadata PDF files and EML are dynamically generated from metadata database. Our new web pages are dynamically generated, so that they can be more easily maintained and updated. The new web site is available on line (beta version), and will soon replace our old web site. Moreore information will be added to our new site over the next year.

In the coming year, we will continue to work on refining our new web site, database, and EML. More information and web search engine will be added to our web site. Metadata and dataset query interfaces and search applications will be our focus in order to improve data management. We will build query interfaces that will enable data searches based on watersheds or treatments, years, and key words in metadata. We will continue to work on our EML to ensure compliance with the Best Practices, and contribute to LTER Metacat. In a word, I am excited for this opportunity to join the IM community and share our information with you.

See you all in Montreal!

LTER Site: Luquillo LTER

Contributor: Eda C. Melendez-Colom (Aug 01, 2005)

Site Byte:

This last year, LUQ IM has been devoted to make progress in two areas: EML translation of all its metadata and local automation of the data manipulation to contribute to the NIS modules, namely, CLIMDB and All-Site Bibliography.

Conversion to EML was divided into two phases. First all the general information about the database's related project, the abstract, methods, and personnel (author, contact, other investigators), among other EML elements were to be converted into EML and harvested into Metacat. A second phase, which will turn LUQ's EML level to a 5th level, will include identities along with their attributes. With the help of Iñigo from the LNO, we completed Phase 1. This puts LUQ's EML into a Level 3 to 4.

LUQ's local metadata database went through a considerable transformation such that when updating or adding EML packages in the future, the transformation will be done automatically. Iñigo developed perl scripts that transformed xml files into EML packages.

The completion of Phase 2 is projected for next year, probably for this time of the year. This phase not only requires the transformation of the metadata, but of the data files themselves. As it turned out many of our data files were not exactly in a database-ready format. Transforming all these data files into a suitable databased format, has delayed this process. We believe that, at the end, LUQ's metadata and data files will be even more suitable for synthesis as well as for EML transformations.

Special attention was given to the conversion of meteorological data into ClimDB. Tables that allows the user to convert regular mm/dd/yyyy-formated dates into the ClimDB dates format (strings with the format "yyymmdd") were developed and the queries within Paradox that transform the dates using these tables were designed also. The rest of the transformation varies for each of the three stations whose data LUQ contributes for ClimDB each requiring different scripts or permanent queries for their transformation. All three have been updated or are in the process of being complete.

SiteDB was updated as well as the All-sites Bibliography (with 617 references as of July 31). Since all LUQ LTER Bibliography is entered in End Notes, the export process of the references is done by End Notes' built in export format, which is accepted by the All-site Bibliographic module. This made LUQ collaboration to the NIS modules complete.

LTER Site: McMurdo Dry Valleys

Contributor: Chris Gardner (Aug 03, 2005)

Site Byte:

This year has been one of transition for MCM in terms of information management. The data management operations were moved from the University of Colorado at Boulder, where a student acted part time as the data manager, to Ohio State, where I will operate full time for the next couple years. We purchased a brand new SunFire V480 server with Solaris 10, and I've been working to get our Oracle database and GIS capabilities operating on the new machine. In addition, I have redesigned the look and feel of the website and hope to go live with it this fall at a new domain: www.mcmlter.org.

We have formed a "data management subgroup," which is comprised of one representative from each of the PIs' research groups. We recently met at the MCM LTER meeting in Portland, OR where we discussed ways to improve the data submission process and streamline our operation. I have also been sifting through our core and legacy data to ensure that we have no gaps in our database, and verifying my findings with this group. Our hope is that the subgroup will make the data management aspect of science more interactive for the MCM researchers, and therefore everyone will take a more active role.

Overall we're making a lot of progress at MCM, and after we have our database up-to-date and new website up, we can start to work on the metadata conversion to EML. Sorry I can't make it to the meeting this year!

LTER Site: Moorea Coral Reef

Contributor: Margaret O'Brien (Aug 23, 2005)

Site Byte:

The Moorea Coral Reef information management system will be housed at the Marine Science Institute of the University of California, Santa Barbara. Currently, the Science Coordinator (Andy Brooks) is filling the role of Information Manager, with consultation from Margaret O'Brien, the IM at our sister site, Santa Barbara Coastal, and Chris Jones of the PISCO project. MCR, SBC and PISCO share several investigators, post docs and staff, and so logically, these project will also share information management software and strategies. MCR will actively start its search for an Information Manager in the fall of 2005 whose skills will complement those of our collaborators..

Fieldwork is conducted at the UC Berkeley Richard B. Gump South Pacific Research Station on Moorea near Lat: -17.5, Lon -149.83. Our researchers are from 4 campuses of the University of California (Santa

Barbara, Davis, Santa Cruz, Davis) plus California State University, Northridge and the University of Hawaii. 2005 was the first active field season for the Moorea Coral Reef site. This season, we installed 3 permanent moorings and several seasonal moorings around the island and in lagoons to investigate coastal oceanographic processes. We also established permanent transects on all 3 sides and initated studies of coral structure and function, fish population dynamics, and of nutrient delivery at the reefs. Data has been delivered to the fileserver in Santa Barbara, and we anticipate insertion of data into the LTER-PISCO Metacat data catalog soon after analysis.

LTER Site: Niwot Ridge LTER

Contributor: Todd M. Ackerman (Aug 02, 2005)

Site Byte:

This past year we have continued to hammer out the conversion of our metadata into the EML format. A few different tactics were attempted and we decided to use 'home cooked' perl scripts. Perl DBI scripts have been developed to allow the text metadata headers of the Niwot datafiles to be updated following the existing protocols at Niwot, then these changes are then loaded into a MSSQL Server database. A cgi script then allows the EML to be harvested from the database (much thanks to Inigo for the harvesting help). Currently the content is somewhere between level 2 and level 3, and some minor additions to the database should allow level 3 content shortly. Our spatial data metadata is the next project to tackle, as well as determining the easiest method of getting the initial metadata into the database.

We have added a few more sites to ClimDB/HydroDB this year, those sites are C1 (sub-alpine, 3022m), D1 (alpine tundra, 3739m), and the hydrological site at Albion (3259m). Success with our wireless program (much thanks to our Climatologist Mark Losleben) has allowed us to now have eight field sites downloading near-real-time data, which has permitted us to add a few more real-time graphs to our website. This has proved very useful in error detection, allowing us to fix problems which previously would have gone unnoticed and lead to months of useless data.

The website received a MUCH needed facelift, bringing our web representation out of the early 90's and into the 2000's (http://culter.colorado.edu/NWT/). The focus was on increased usability and content, and of course appearance.

We have continued development on our Data Entry Form System (DEFS) which has been designed in VB .NET to greatly improve data entry of our field data sheets. This is a great improvement over our previous DOS-based key/re-key entry system, where data entry forms are designed to

mimic the data collection forms.

LTER Site: North Temperate Lakes

Contributor: Barbara Benson (Aug 02, 2005)

Site Byte:

from Barbara Benson, Dave Balsiger, Jonathan Chipman, and Paul Hanson

Considerable work has been done in implementing the Ecological Metadata Language (EML) standard at NTL. The EML elements included in the NTL EML documents were expanded; particularly notable is the addition of taxonomic coverage. In April 2005 NTL LTER IM staff harvested EML documents for most of the NTL core data sets into the central metadata catalog, Metacat, for the LTER Network. These harvested EML documents are valid EML and describe identification, discovery, evaluation, access, and integration information. We are in the process of developing the metadata content further to be in full compliance with the EML Best Practices document developed by the LTER Information Managers.

The NTL LTER website underwent a major redesign during the past year. With the help of a professional graphics/web designer the website organization, appearance, navigation and usefulness were significantly enhanced. Website content was brought up to date.

Our development in the area of sensor networks has continued with the addition of another instrumented buoy at NTL and our leadership in the Global Lake Ecological Observatory Network (GLEON), a grassroots network of limnologists, information technology experts, and engineers who are coordinating the construction of a scalable, persistent network of lake ecology observatories. In March 2005 an international workshop of researchers studying lakes and coral reefs was held in San Diego to explore building capacity using sensor networks and linking infrastructure to share data. This year personnel from NTL visited the instrumented buoy on Yuan Yang Lake in Taiwan to help upgrade the instrumentation and interact with Taiwanese scientists. Within the next year monitoring buoys on lakes in New Zealand, Israel and Finland should be added to the global network. Lakes in mainland China, South Korea, Japan and Australia may be added to GLEON in the future. NTL is collaborating through an NSF grant with computer scientists at UCSD, SUNY-Binghamton, and Indiana University to solve problems that limit the extensibility and scalability of data-generating sensor networks: (1) automating instrument management and the updating of data flows from sensors to publicly-accessible biological databases and (2) developing a suite of new algorithms and software for detection (real-time) of events

based on data from sensors and databases, with applications to classification of signals as deriving from biological or physical events or to sensor failure, allowing rapid response.

Programmers at the Center for Limnology are developing software that provides sample and data management for the Center for Limnology Chemistry Lab (the program is called ChemLab). This project was initiated because the overload on the existing chemistry lab data management was leading to unacceptable sample processing backlogs. Tracking status of bottles and tests using pen and paper is inefficient and time consuming, and there was duplication of effort with the field crew in labeling and tracking bottle metadata. The ChemLab software stores bottle metadata, bottle/test status and analytical results in a database accessible to both the field crew and the chem lab technicians.

The spatial data catalog was improved and upgraded in several significant ways. The existing “static” version of the Spatial Data Catalog was redesigned as part of the overall website redesign. All the spatial data sets were converted to a common file format and map coordinate system, and uniform metadata were created. Several new data sets were added, most notably demographic spatial data. Finally, we developed a new, alternate method for managing the spatial data, which permits direct, dynamic access to the data via an Oracle database with an ArcSDE (Spatial Data Engine) interface. This represents an improvement over the “static” spatial data catalog by (a) facilitating access to data and metadata; (b) ensuring that users are working with the current version of all data; (c) enabling development of cross-site or network-level spatial data applications; and (d) facilitating development of web-based mapping applications. Ongoing spatial data management activities include work on such web-based mapping systems, the addition of new data sets to the catalog, and testing methods for converting the XML-format spatial metadata to valid EML.

LTER Site: Palmer Station

Contributor: Karen Baker (Aug 27, 2005)

Site Byte:

PAL Information Management efforts focused on implementing a second-generation system designed over the last few years following research into and experience with site practices, federated network criteria, and metadata standards' requirements. Study of the infrastructuring concept along with last year's hardware implementations created a contemporary infrastructure within our local UCSD/SIO Integrative Oceanography Department so that the PAL data project could be migrated to a new storage system. Update and reorganization of the decade old file structure and its content is ongoing. The informatics team with designated

information manager Karen Baker broadened to include Shaun Haber and Mason Kortz as web and database designers, Lynn Yarmey as dictionary and metadata analyst, and Jerry Wanetick as computational center director and systems administrator.

Amid a year of multiple transitions, the site highest priority data management task was development of a long-term, extensible metadata strategy, bridging from decade old text forms to a relational database approach. From this effort emerged the recognition of a need to develop unit and attribute dictionaries. In considering the site information system holistically, the design and implementation of a personnel directory represents the initial module an important element to an integrated approach. Along with these activities, a Palmer web site redesign includes a three tier template, stylesheets and update of dynamic elements such as the photo gallery, glossary, and sampling grid program under the new architecture.

Collaborative local activities included coordination with the colocated LTER California Current Ecosystem (CCE) site and California Cooperative Oceanographic
Fisheries Investigations (CalCOFI) programs along with the Southern California Coastal Ocean Observing System (SCCOOS) program. The design and development of an Ocean Informatics Environment is ongoing, providing a comprehensive conceptual framework for all informatics activities. Making use of a design studio approach along with strategic design teams and working groups contributed to community efforts such as a joint data acquisition schema and promoted a shared understanding across multiple data types. Interaction with the Comparative Interoperability Project is providing opportunities to consider and articulate how IM work is carried out within different communities (Ribes et al, 2005; Baker et al, 2005; Millerand et al, 2005) and prompted development of two local reading groups. In additional, PAL and CCE contributed the notion of 'social informatics' to the first meeting of LTER social scientists in August of this year.

LTER Network activities included participation with the Dictionary Process Unit Repository design team that created a prototype web application in time for the annual Information Management meeting. Further, a Community Process Working Group provided a mechanism for exploring lessons learned in the design, development, implementation, and enactment of community standards.

LTER Site: Plum Island Ecosystem

Contributor: Hap Garritt (Aug 03, 2005)

Site Byte:

The Plum Island Ecosystems (PIE) LTER has recently, through the help of the LTER Network Office (Inigo San Gil and Duane Costa) and Jim Laundre ARC successfully converted 119 non-spatial legacy metadata files to ~ Level 2.5 EML including harvest into the LNO Metacat. We are continuing to make progress in standardizing a metadata document for the three different spatial data softwares (ESRI ARC, Clark University IDRISI and University of New Hampshire GHAAS-RGIS in use at PIE which will be followed up by conversion to EML.

In June 2005, PIE LTER hosted a workshop for developing plans for upgrading the Marshview Farm field facility into the Plum Island Coastal Research Facility. Attendees included PIE research personnel and also outside researchers who have had experience in managing field research facilities. Working groups met to discuss facility and infrastructure needs in the program areas of Research, Education and Outreach with the intent of developing new facilities that would complement existing available facilities in the area. The recommendations from the workshop will be incorporated into a facilities upgrade proposal in the near future

---

LTER Site: Santa Barbara Coastal

Contributor: Margaret O'Brien (Sep 12, 2005)

Site Byte:

This year as in previous, Santa Barbara Costal LTER's Information Management system collaborated with other local coastal and oceanographic programs with whom we also share collection and analysis methods and personel. Our data catalog is shared with PISCO (administered by Chris Jones and Jordan Morris) and we will be soon joined by the Moorea Coral Reef LTER site (whose Information Manager is yet to be hired). SBC's data needs are managed by Margaret O'Brien with reprentatives from its major research groups who provide data and metadata and receive introductory instruction to data storage methods (e. g., EML) as needed. Also during 2005, the LTER-PISCO IM office moved into a new building on the UCSB campus, putting us on the same floor as the NCEAS Ecoinformatics programmers. We are enjoying fruitful interactions with this group, also.

Data Catalog: During the past year, the contents of our Metacat data catalog was updated from version EML2.0beta6 to EML2.0.1. About 70% of core data is currently included and is at approximately level 4. For browsing, we have created canned queries using a thesaurus of research categories, and keyword and/or site queries are planned. The remaining core data will be added to the catalog by the end of the calendar year. In general, SBC has chosen the strategy of bringing "fewer datasets to a higher level". This is efficient because most of SBC's data goes (or will go) directly into metacat from collection/analysis, with

multiple datasets making use of a common eml template with full attribute discriptions. Some of our current data packages will serve as templates for new packages. Currently, some data packages contain several data tables if these shared spatial, time or protocol coverage, however some of these will be separated to allow finer-grained searches. During the summer 2005, our metacat began replicating to the KNB network. Chris Jones has led an effort to redesign XSL stylesheets for displaying EML content in a modular form in HTML pages, using Cascading Style Sheets to control look and feel. This project is well underway, but still requires considerable work. Two implementations can be viewed at the data links below.

In addition to the data catalog, we have also been investigating the uses of EML for handling other information. We have recently inplemented a variant of EML in Metacat for our site's bibliography. We have found EML to be quite well suited to our needs, providing mechanisms for a separate listing of in-press pubs, fairly quick displays and filtering, document distribution, and possibly, a link to an associated dataset. This project is ongoing, and may result in some products that will be useful to the network as a whole.

Quite soon, we are anticipating upgrades in both our server hardware and Metacat software. In late August, we plan to migrate our data server and catalog to an IBM x336 with a 2.4tb raid array. SBC is already preparing for its 2006 renewal, and so a good deal of IM time has gone into assisting researchers with their synthesis projects. For the next several months, most of the IM projects will have data access, integration and synthesis as their promary goal.

Further information:
data catalogs with xsl/css style:
http://data.piscoweb.org
http://sbc.lternet.edu/data

bibliographic db (under construction):
http://sbc.lternet.edu/catalog/style/skins/sbclter/sbcPubsDisplay.jsp

---

LTER Site: Sevilleta LTER

Contributor: Kristin Vanderbilt (Jul 29, 2005)

Site Byte:

The Sevilleta LTER Information Management group continues its conversion of our text-file based information management system into MySQL. More data entry and QA/QC tools were developed during the past year that capitalize on the relational database. A particular coup for the Sevilleta IM, Kristin Vanderbilt, was getting the SEV publications

database moved into MySQL and a web interface created that allows PIs to enter their own publication data.

The Sevilleta IM team has embraced wikis as a mechanism for posting FAQs about systems administration, data entry instructions, protocols, and data processing instructions. The wikis have been integrated into our PostNuke-based web page.

We continue to translate our legacy semi-structured text metadata into EML. It's a time-consuming process, because for some of the data we have used Morpho for the conversion. Inigo San Gil from LNO has been working with us recently to develop a script that will do the conversion for much of our data. The output of the script will include attribute metadata, so we will be well on our way to Level 3 compliance as described by the EML Best Practices document. By the end of the year, with the help of Inigo, we expect to have most of our metadata available as EML. We successfully harvested some EML documents into metacat on July 28, 2005.

Kristin continues to be involved in outreach projects. She teaches portions of ecoinformatics training classes for personnel from OBFS and for junior faculty and postdocs who participate in a training course sponsored by the Science Environment for Ecological Knowledge project (http://seek.ecoinformatics.org). She is also involved in a research project with Judy Cushing from Evergreen State College, wherein templates for ecological databases are being designed for a pilot project focused on integrating grassland LTER NPP data. She also participated in a panel at the SSDBM meeting in June 2005 entitled "NSF Long Term Ecological Research Sites: Praxis et Theoria—LTER Information Management and CS Research." Kristin also presented a poster entitled "Ecoinformatics Training: Toward Data Sharing and Collaborative Research" at the Workshop for Enhancing Collaborative Research on the Environment in Sub-Saharan Africa in Arlington, VA in January 2005.

LTER Site: Shortgrass Steppe

Contributor: Nicole Kaplan (Jul 21, 2005)

Site Byte:

The Shortgrass Steppe LTER Information Management (IM) team is working on projects to improve support for local and network science, and access to more integrated, metadata, data and other information. The team currently consists of Nicole Kaplan, who works closely with Bob Flynn, GIS and IT manager. A position on the team has been created to support database and web development, which is currrently being advertised for a comupter science student. It is important to balance our commitment to site support, Network initiatives and outside

ecoinformatics projects. The SGS Principal Investigators are discussing the interpretation and application of newly adopted LTER Network data access and use policies. Nicole is involved with the Network IM and broader ecoinformatics communities as a member of LTER IMexec and Website Design Working Group Leader.

The SGS Information Management team has plans to improve on-line searching capabilities for data and metadata by developing new web site tools. They have recognized the need for better integration of related data sets, spatial and non-spatial data, publications and other research information. Recommendations from the LTER Network Wesite Design Working Group will be considered when implementating the second generation SGS web site (Kaplan 2005). Our field staff are in the process of collecting Global Positioning System coordinates to complete our spatial coverage of most legacy and all current long-term and short-term data stes. We are making progress toward integrating our non-spatial and spatial data sets, to enhance management of study impacts on field sites and support data discovery with level 2 Ecological Metadata Lanaguage from the SGS database.

Metadata in the SGS Information Management System vary in richness, since fifteen percent of our data sets are legacy data sets from the USFS and IBP. Because of this, information required to meet the new LTER standard metadata content of EML may not be available. Metadata for more recent and current SGS data sets can be submitted directly to the RDBMS by students and PIs via web-based forms. We recognize the importance of maintaining robust metadata to ensure the usability of data in the future and are making efforts to conform our metadata tables in the RDBMS to EML according to the EML Best Practices document. We have a strategy to bring metadata in the RDBMS to Level 2 EML by exporting metadata content in XML (Extensible Mark-up Language) and converting the XML to EML with XSLT (Extensible Stylesheet Language Transformation) conversion scripts. Experts at the LTER Network Office have contributed to our efforts by providing licenses for software tools, example code, and tools for harvesting SGS metadata to a Metacat, a remote ecological metadata catalog. The SGS has also contributed to a community model metadata management system in RDBMS that is being developed and implemented at various LTER sites by the LTER Network office. Lastly, a GIS EML tool developed at the CAP (Central Arizona Phoenix) LTER site was tested on SGS GIS metadata to generate EML.

The IM Team is developing a suite of programs to improve QAQC practices at SGS, called the Matrix. The Matrix currently checks and formats meteorological data for submission to CLIMdb (http://www.fsl.orst.edu/climdb/) and is being expanded to support data tables produced by floral dynamics research, which contains over sixty percent of our studies. The IM team will continue to work with researchers to develop tools to more efficiently process, quality check and publish their data with high integrity.

Nicole continues to participate in ecoinformatics community projects,

such as the Canopy Databank Project (http://canopy.evergreen.edu/bcd/home.asp) at The Evergreen State College. LTER IMs are contributing to the development of templates for data entry, analysis and synthesis of aboveground net primary production data collected at distinct grassland sites. Synthesis of nutrient enrichment data is being demonstrated by the Science Environment for Ecological Knowledge (http://seek.ecoinformatics.org/) project and includes SGS data and input for constructing ontologies of aboveground grassland vegetation measurements.

References

Kaplan, N.E., C. Gries, and E.C. Meléndez-Colom (2005) Evaluating First Generation LTER Site Web Sites: Assessing our audience, meeting their needs, and making recommendations for the future. In DataBits: An electronic newsletter for Information Managers, Spring 2005 Issue.

LTER Site: Virginia Coast Reserve

Contributor: John Porter (Jul 12, 2005)

Site Byte:

It has been another busy year at the VCR/LTER, with a substantial revamping of the site web page (http://www.vcrlter.virginia.edu) and development of a program that automatically generates EML metadata from our existing metadatabase. We have also been involved with several outreach and service functions, including web support for the Statistical Ecology (http://www.esa.org/stat-ecol) and Long-Term Studies (http://www.esa.org/longterm) Sections of the Ecological Society of America and the Mid-Atlantic Region Ecological Observatory (http://mareo.org) and work on databases with the Taiwan Ecological Research Network (TERN). Additionally, we have been expanding our wireless networking capabilities to reach new field instrumentation.

We have been making extensive use of the PostNuke Content Mangement System (CMS) for developing special-purpose web pages with generally good results. As an open-source product aimed at "community" web sites, PostNuke provides many useful functions, but is not without limitations, the most notable of which is the inability to make menu bars page-specific. However, on balance it has worked surprisingly well.

Ecological Metadata Language (EML) production is now automated with a PERL program "make_eml.pl" that queries metadata from a relational database system using the perl DBI functions (allowing movement to another brand of database at a later time). Although we are able to generate more-or-less full EML metadata, there are areas where we need

to alter our underlying system to accomodate EML needs. These include adding types to keywords (currently everything is listed as "thematic"), making measurement units more consistent, and improving information on taxonomic coverage. Aside from its length (due to the number of EML elements), the program was relatively easy to write - consisting of repeated "query -> create tag" sections. The biggest problem has been dealing with free text, where users make inconsistent use of paragraph markers in the text, thus potentially unbalancing XML tags and causing validation problems.

During the spring of 2005 we hosted Meei-ru Jeng of the Taiwan Forestry Reseach Institute for three months. During that time she worked on learning PERL, PHP, Mapserver, PostNuke and how to make these programs interact with MySQL databases. To hone her PERL skills and to help us reconcile our literature and personnel databases, she developed a program that creates an EndNote Import file wherein names and intiials of VCR researchers have full names substituted. These interactions continued this summer with visits by John Porter to Taiwan and China. Currently plans are being made for additional extended visits by Taiwanese information managers to US LTER sites.

Following up on successful wireless networking workshops at the LTER All-Sci meeting in 2003 and at the Ecological Society of America meeting in 2004, we wrote a paper for BioScience that was the cover story in the July 2005 issue. This is acompanied by yet-another-postnuke-web-site: http://wireless.vcrlter.virginia.edu. We hope that LTER information managers will be active contributors to that "community" web site as we gain additional experience with wireless networking.

Home

Agenda

Participants

Resources:

Site Bytes

References

Presentations

Working Groups

Links:

LTER Home

LTER IM Page

IM2004 Report

IM2003 Report

Pre-meeting Reference Materials

Links to reference material intended to inform participants about general topics for discussion at the meeting. Everyone should familiarize themselves with these documents prior to attending.

Briefing Documents -

» EML Best Practices document, example xml files and xml templates (version 1.0, October 2004) -- *contributed by Wade Sheldon (GCE)*

» "Incorporating Semantics in Scientific Workflow Authoring" is a brief paper published from SSDBM 2005 that describes some aspects of the development, use, and capabilities of ontologies within the SEEK and Kepler frameworks. It may be good background for the ontology breakout groups. -- *contributed by Matt Jones (Guest)*

» LTER Network Planning Grant Wiki Site --- background information on the planning grant activities and working group reports. Of special interest are the documents for the Cyberinfrastructure working group. -- *contributed by Barbara Benson (NTL)*

» Cyberinfrastructure Supplement proposal for the LTER Network Planning Grant --- funded grant proposal to conduct the cyberinfrastructure planning to support the research agenda being developed by the LTER Network Planning grant -- *contributed by Barbara Benson (NTL)*

General Reference -

» LTER IM Committee web page -- *contributed by Wade Sheldon (GCE)*

» LTER Network Data Access Policy, Data Access Requirements, and General Data Use Agreement -- *contributed by Mark Servilla (LNO)*

» LTER Network Information System Strategic Plan v2.8.1 -- *contributed by Mark Servilla (LNO)*

» Review Criteria for LTER Information Management Systems v1.0 -- *contributed by Emery Boose (HFR)*

Home
Agenda
Participants

Resources:
Site Bytes
References
Presentations
Working Groups

Links:
LTER Home
LTER IM Page
IM2004 Report
IM2003 Report

Presentations and Reports

Links to presentation and report files contributed following the meeting intended for follow-up discussions and compiling the meeting report.

Presentations -

» McCartney presentation on Data Access Policy Revision. -- *contributed by Peter McCartney (CAP)*

» Canopy Databank Project presentation -- *contributed by Judy Cushing (Guest)*

» Metacat Advanced Query Interface -- *contributed by Mark Servilla (LNO)*

» LTER Custom Unit Registry -- *contributed by Mark Servilla (LNO)*

» System Internationale (SI) and STMML Background -- *contributed by Mark Servilla (LNO)*

» LTER Grid Pilot Study -- *contributed by Mark Servilla (LNO)*

» EML status of the LTER sites -- *contributed by inigo san gil (LNO)*

» Network Information System Strategic Plan -- *contributed by James Brunt (LNO)*

» Network Databases Status Report -- *contributed by James Brunt (LNO)*

» California Current Ecosystem (CCE), new site presentation -- *contributed by Lynn Yarmey (CCE)*

» Moorea Coral Reef (MCR), new site presentation -- *contributed by Margaret O'Brien (MCR)*

» General content summary of ClimDB/HydroDB -- *contributed by Don Henshaw (AND)*

» Sensor Networks -- *contributed by Barbara Benson (NTL)*

» CI in LTER Network Planning Grant -- *contributed by Barbara Benson (NWT)*

Working Group Reports -

» Report of Attribute working group Saturday Morning -- *contributed by*

Peter McCartney (CAP)

» Working Group 1 Report: SiteDB -- *contributed by Nicole Kaplan (SGS)*

» Working Group 4 Report: Recommendations for Web Sites -- *contributed by Nicole Kaplan (SGS)*

» Working Group 7 Report: Community Process -- *contributed by Karen Baker (PAL)*

» Site Innovations Working Group Report -- *contributed by John Porter (VCR)*

» Query Interface/Controlled Vocabulary Working Group Report -- *contributed by John Porter (VCR)*

» Summary of working group meeting for planning IM participation in 2006 All Scientists Meeting. -- *contributed by Jonathan Walsh (BES)*

Home

Agenda

Participants

Resources:

Site Bytes

References

Presentations

Working Groups

Links:

LTER Home

LTER IM Page

IM2004 Report

IM2003 Report

Working Groups

Links to descriptions, briefing documents and other resources for working groups that will be convened at the meeting. Links and files can also be contributed during and after the meeting to support ongoing collaborations.

Group 1: Web Access to Information: SiteDB -

» Working Group 1 Description of the issues and goals we will discuss regarding SiteDB. -- *contributed by Nicole Kaplan (SGS)*

» Web address for SiteDB -- *contributed by Nicole Kaplan (SGS)*

» An overview of the SiteDb, including its attributes -- *contributed by Eda C. Melendez-Colom (LUQ)*

» Results of a survey done on the population of the general information section of SiteDB by the LTER sites -- *contributed by Eda C. Melendez-Colom (LUQ)*

» Working group 1 (siteDB) Report -- *contributed by Nicole Kaplan (SGS)*

Group 2: Design of Query Interface to LTER Catalog: Controlled Vocabularies -

» Background on working group -- *contributed by John Porter and Jonathan Walsh (VCR)*

Group 3: Exploring Ontologies: Background -

» Home Page for EPA Storet database; a widely used standard for encoding chemical measurments. -- *contributed by Peter McCartney (CAP)*

» GEON project: Follow link to myGEON to see example of an ontology-driven query and integration system. -- *contributed by Peter McCartney (CAP)*

» Working group summary -- *contributed by Peter McCartney (CAP)*

» Discussion notes from Friday work group sessions -- *contributed by Peter McCartney (CAP)*

» SWEET Ontologies from JPL -- *contributed by Peter McCartney (CAP)*

Group 4: Web Site Design Recommendations -

» Working group 4 description of challenges, successes and recommendations for developing new web sites at LTER sites. -- *contributed by Nicole Kaplan (SGS)*

» Web user survey used to identify main user groups of LTER web sites, contributors to web site development, and successful web tools in place. -- *contributed by Nicole Kaplan (SGS)*

» Working Group 4 (Web Site Recommendations) Report -- *contributed by Nicole Kaplan (SGS)*

Group 5: Unit Dictionary -

» Dictionary Process: Unit Registry Prototype - Working Group Description -- *contributed by Karen Baker (PAL)*

» Dictionary Process: Unit Registry Prototype - Working Group Report -- *contributed by Karen Baker (PAL)*

Group 6: Attribute Ontologies -

» Overview slides for Attribute Ontology working group, Friday Morning -- *contributed by Peter McCartney (CAP)*

Group 7: Community Process and Standards Implementation -

» Community Process and Standards Implementation: Working Group Description -- *contributed by Barbara Benson (NTL)*

Group 8: Joint: Ontology, Unit Dictionary, Controlled Vocabulary Groups -

» Notes from Joint meeting Saturday Morning -- *contributed by Peter McCartney (CAP)*

Group 9: All Scientist Meeting 2006 Planning -

» Working group summary -- *contributed by Ken Ramsey (JRN)*

» 2003 LTER All-Scientist Meeting home page -- *contributed by Ken Ramsey (JRN)*

Incorporating Semantics in Scientific Workflow Authoring

Chad Berkley, Shawn Bowers, Matthew B. Jones, Bertram Ludäscher, Mark Schildhauer, Jing Tao
{berkley, jones, schild, tao}@nceas.ucsb.edu, {bowers, ludaescher}@ucdavis.edu

Abstract. *The tools used to analyze scientific data are often distinct from those used to archive, retrieve, and query data. A scientific workflow environment, however, allows one to seamlessly combine these functions within the same application. This increase in capability is accompanied by an increase in complexity, especially in workflow tools like Kepler, which target multiple science domains including ecology, geology, oceanography, physics, and biology. To overcome this complexity, we have developed semantically-driven user-interface components that are customized at run-time using domain-specific ontologies. One such subsystem in Kepler uses domain-specific ontologies to customize the presentation of analytical components and data for use by scientists building workflows. Kepler also provides for semantically-enabled queries for components, which can significantly increase efficiency in workflow authoring tasks. In this demonstration, we show how ontologies can be used for user-interface customization and more. In particular, we show our recent ontology-driven extensions for workflow authoring in Kepler. These extensions include our advances in: (1) automating data-integration and service-composition tasks, (2) the use of semantic annotations to verify that workflows are semantically meaningful, and (3) the ability to search for contextually relevant components and data sets* in situ, *i.e., as a user is designing a scientific workflow.*

1. Introduction

Scientific workflow systems have traditionally been stand-alone applications designed for a specific domain. For example, physicists, geologists, ecologists, and oceanographers typically use their own applications (e.g., a set of "MATLAB" scripts) for creating and executing scientific workflows. The Science Environment for Ecological Knowledge (SEEK) [SEEK] project is developing a powerful, cross-domain scientific-workflow authoring environment that allows scientists to design and execute novel workflows. The need for such a tool has been recognized in other scientific domains, and so SEEK has teamed up with several other projects, including GEON [GEON], SDM [SDM], EOL [EOL] and ROADNet [ROADNET] to produce *Kepler* [KEPLER].

Scientific workflow systems such as Kepler provide scientists with a number of benefits. In particular, they provide an integrated environment in which scientists can design, communicate, and execute their analytical processes. They typically incorporate a variety of functions for end-to-end workflow execution and management, including data query, retrieval, and archiving tools. And, they provide a mechanism to help scientists recreate previous analyses (thus allowing workflows to serve as a form of metadata) and provide an opportunity for workflows (and data) to be reused to form novel and extended analyses.

A major challenge for Kepler is to effectively support users from different scientific disciplines, while maintaining both generic support for scientific workflows and enabling cross-domain data and workflow reuse. Instead of creating complex interfaces and tools for each domain, we desire the capability to provide domain-specific customization. We believe that ontologies can be used not only to formalize domain knowledge, but also to support creation of customized user interfaces, thus facilitating cross-domain interaction.

As part of SEEK (and in collaboration with the other projects previously noted), we are actively engaging scientists to develop ontologies, with the goal of having a rich repository of domain-specific terminologies and cross-linkages among them. Along with this effort, we are also developing a suite of ontology-based tools [BLL04, BL04, BTWL04] to allow scientists to more easily browse, query, integrate, and compose relevant cross-discipline datasets and services. This demonstration will highlight these ontology-enabled tools and their implementation within Kepler.

2. Scientific Workflows and Kepler

A scientific workflow is an executable description of a scientific process. In particular, a scientific workflow records each inline process required to take input data and produce a meaningful output product. Scientific workflows are similar to business-process workflows but have several properties uncommon to the business environment. For example, scientific workflows often operate on large, complex, and heterogeneous data. They can be computationally intensive, and can produce complex derived data products that may be archived for use in re-parameterized runs or other workflows. Moreover, unlike business workflows that are often event-flow driven, scientific workflows are generally data-flow driven (i.e., execution is based on the flow of data as opposed to triggered events).

In Kepler, scientific workflows bring together data and services, possibly created by groups or individuals unknown to each other. Moreover, the workflow applications written in Kepler encompass a wide variety of scientific domains, sub-domains, and specialties. By making these data and services broadly accessible and comprehensible way, Kepler facilitates cross-domain investigations and interdisciplinary research.

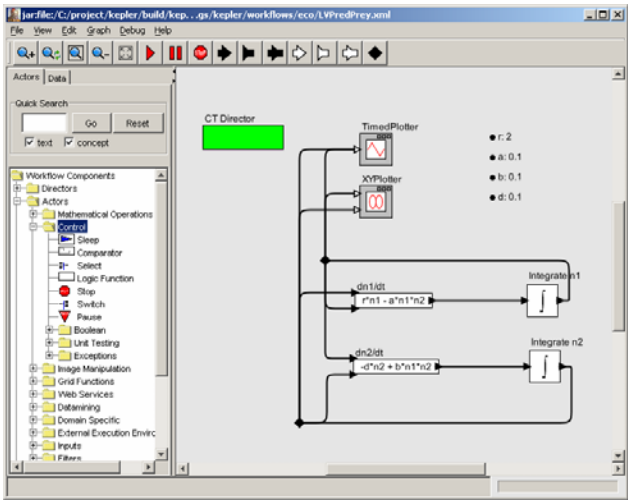


Figure 1. The Kepler scientific workflow environment.

Within Kepler, scientific workflows are authored in a graphical, drag-and-drop manner. Services contain typed ports that can be connected to other services or data sources. Ports can have simple atomic types such as *integer* and *string* as well as more complex structures, including arbitrarily nested array and record types. As a workflow is executing, data passes between ports via tokens that can be readily manipulated to meet the differing syntactic needs of other services. Data produced by a scientific workflow can be displayed graphically at run time, or written to disk for later use.

An example scientific workflow within Kepler is shown in Figure 1. The panel on the left is the "library" where components are categorized and can be searched by a user. When a component is needed on the canvas (the panel on the right), it is dragged from the library onto the canvas where it can then be configured and have its ports connected to other components. The green box controls the timing and flow of the model and can also be selected and drug from the library.

3. Conceptual Challenges in Scientific Workflows

Because Kepler is a powerful and flexible workflow system with a diverse set of users, a number of conceptual-modeling challenges arise. Our goal is to allow users with different backgrounds and varying levels of computing expertise to create new scientific workflows with a minimum amount of difficulty. We highlight below the main difficulties we wish to address.

Supporting high-level conceptual models. Most scientists have a high-level conceptual model of their workflows. If asked, a scientist can typically write down the steps involved in taking raw data and producing their desired output fairly quickly. However, when this conceptual model becomes formalized into an executable scientific workflow, a large number of low-level technical details arise. Details such as file access, network protocols, dataset schemas, service input and output typing, execution models (e.g., tuple-at-a-time versus table-at-a-time dataflow), and configuration parameters tend to obscure the high-level conceptual model of the workflow, making it hard to compare it with existing workflows and reuse it in new settings. We would like to effectively capture the high-level aspects of a workflow, while also preserving but often hiding the underlying technical details.

Basic contextual metadata. A general lack of contextual metadata with respect to data and services is problematic for users (e.g., those who are trying to find new and relevant datasets and services). As an example, a service titled "interpolator" might give one the impression that it provides a generic interpolatation operation over arbitrary datasets when in fact, the service was written to interpolate spatial grid data. Additionally, the same component could simply have been named "int," obscuring the functionality of the service even though those familiar with the particular workflow know that "int" means "spatial data grid interpolate". We face the challenge of making these services generically comprehensible and accessible.

Schema and service-type semantics. Scientific data integration can be a complex and time-consuming process. Scientific data is highly heterogeneous, laden with structural, schematic, and semantic differences. Today, scientific-data integration is typically performed by hand and requires significant "meta" information. Service composition similarly requires considerable contextual information describing structure (to manage heterogeneity in input and output types) and semantics (the kind of objects consumed or produced by a service).

4. Using Semantics in Workflow Authoring

Robust metadata is required to meet the challenges involved in enabling domain scientists to create, run, and share scientific workflows. Several communities continue to have grass-roots organizations that deal with the collection and storage of syntactic metadata. The Knowledge Network for Biocomplexity [KNB] serves the ecological community with the Ecological Metadata Language (EML) [EML] and associated metadata repositories [JBBS01]. Other relevant metadata standards for Kepler include FGDC [FGDC] and Dublin Core [DC], to name a few.

While standards such as EML may provide some support for semantic metadata (e.g., using a "keyword" field), this information is typically not sufficiently formalized for general use in an automated environment. Most current metadata standards for services also fail to include such formal semantics, including the Web-Service

Description Language (WSDL) [WSDL] and the Modeling Markup Language (MoML) [MOML].[1]

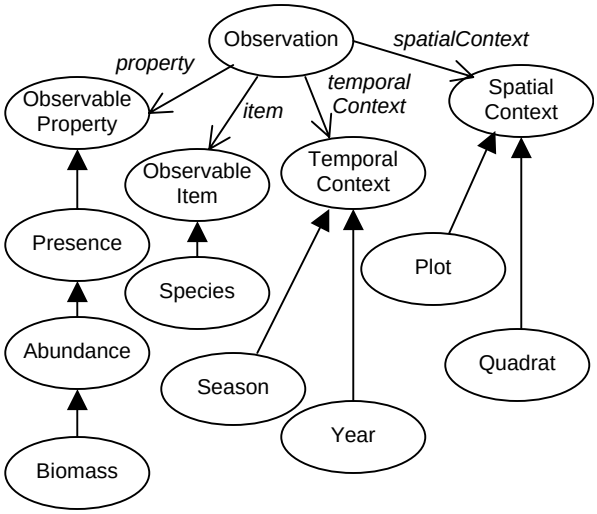


Figure 2: A simplified SEEK ontology.

Kepler has adopted the OWL web ontology language [OWL] (more specifically, OWL-DL) as the primary language for capturing domain-specific terminologies used in semantic metadata. Our approach is to leverage OWL-DL ontologies and *semantic annotations* (described below) of data and services within Kepler to capture rich and possibly complex semantic metadata. A fragment of the SEEK "measurement" ontology is shown graphically in Figure 2.

To address the conceptual challenges discussed in the previous section, we have developed the following features, which we propose to demonstrate. Each of these features leverages the domain ontologies being developed within SEEK and the other Kepler projects.

Support for detailed semantic annotations. Kepler is designed to provide users with the ability to semantically register [BLL04] their dataset schemas and services (and their corresponding input and output types) using *semantic annotations.* Figure 3 gives a set of semantic annotations for the *biom* dataset containing species biomass observations. A semantic annotation defines a relationship between a service or dataset and terms in an ontology. Intuitively, semantic annotations define the "semantic type" of the resource (shown by the statements on the left of the arrows in Figure 3), and link portions of the semantic type to portions of the resource (shown on the right of the arrow in Figure 2). For example, the first annotation in Figure 3 states that tuples in the *biom* dataset denote *Observation* instances from the ontology (in Figure 2). Similarly, the second annotation states that a year value within a tuple denotes the corresponding observation's temporal context and is an instance of the *Year* ontology concept. The semantic annotation language is designed for use at different "granularities," e.g., from selecting a single concept and assigning it to a service, to prescribing a complex ontology instantiation and assigning individual structures within it to particular data values within a dataset (such as in figure 3).[2]

Dataset Schema:
biom(yr, seas, plt, qd, spp, bm)

Semantic Annotations:

x:biom	==>	*x*:Observation
x:biom[yr=*y*]	==>	*x*[*temporalContext*=y:Year]
x:biom[seas=*s*], *s*='W'	==>	*x*[*temporalContext*=s:Winter]
x:biom[seas=*s*], *s*='S'	==>	*x*[*temporalContext*=s:Spring]
x:biom[seas=*s*], *s*='F'	==>	*x*[*temporalContext*=s:Fall]
x:biom[plt=*p*]	==>	*x*[*spatialContext*=*p*:Plot]
x:biom[qd=*q*]	==>	*x*[*spatialContext*=*q*:Quadrat]
x:biom[spp=*s*]	==>	*x*[*item*=*s*:Species]
x:biom[bm=*b*]	==>	*x*[*property*=*b*:Biomass]

Figure 3: Example semantic annotations.

Workflow-component classification and browsing. Kepler leverages semantic annotations to provide customizable access to datasets and services. As shown in Figure 1, the panel on the left displays hierarchically arranged concepts taken from a user-selected ontology, and automatically places services within the hierarchy. This feature provides Kepler users the ability to: 1) select and configure the classification ontology, 2) view the hierarchically arranged ontology (which is computed using a description-logic classifier), and 3) see services classified according to the concept hierarchy (by matching these up through their semantic annotations). In this way, users can easily customize Kepler service presentation (similarly for datasets), and provide ontology-based browsing of data and services.

Semantic scientific-workflow analysis. Given a workflow of interconnected actors, Kepler statically checks (i.e., at design time) whether two connected services (or data sources) are "semantically compatible" based on their semantic annotations, and notifies the user when a connection is not considered semantically well typed. This capability directly assists a user with the workflow creation process.

Ontology-directed scientific-workflow design. As large repositories of workflow components become available,

[1] An exception is the proposed OWL-S [OWL], which provides a "heavy-weight" language for defining the semantics of services.

[2] Semantic annotations in Kepler differ from other approaches by providing rich semantic descriptions that can be "superimposed" over structural types and schemas, allowing explicit connections between substructures and semantic types.

finding relevant resources becomes more difficult. Given a workflow service on the Kepler canvas (the right panel of Figure 1), a user can search for all "semantically compatible" resources (either datasets or services) that can be connected to the input (or output) of the service. This search can also be restricted to return resources that are both semantically and structurally compatible (using Kepler's type system).

(Semi-)Automated Integration and Composition. Scientists often reuse existing workflow components to construct new models. Such components are more often than not structurally incompatible, even though they may be semantically compatible. Our goal is to exploit semantic annotations to derive structural correspondences between input and output data types [BL04]. These correspondences often contain enough information to derive the desired data transformations, allowing scientists to state the desired component connection instead of the low-level details of how those connections should be made. Similarly, multiple datasets must often be combined (i.e., merged or integrated) to be useful as input to a workflow. In this demonstration, we will also show our recent developments for assisting Kepler users in the process of data integration [BTWL04] and service composition, leveraging semantic annotations and domain-specific ontologies.

5. Conclusions and Future Work

In our poster we will show our recent ontology-driven extensions to Kepler for workflow authoring. These extensions include (1) our advances in automating data-integration and service-composition tasks, (2) the use of semantic annotations to verify that workflows are semantically meaningful, and (3) the ability to search for contextually relevant components and data sets in situ, i.e., as a user is designing a scientific workflow. The utility of these extensions will be shown within the context of developing species biodiversity analyses within Kepler.

Acknowledgments

Kepler is based upon the Ptolemy II code base. Kepler includes contributors from SEEK [SEEK], SDM Center-SPA [SPA], Ptolemy II [PTOLEMY] and Geon [GEON]. Work supported in part by NSF ITRs 0225676 (SEEK), 0225673 (GEON) and DOE Grant DE-FC02-01ER25486 (SDM)..

References

[BLL04] Bowers S., K. Lin, and B. Ludäscher, 2004. On integrating scientific resources through semantic registration, In Proc. of SSDBM, pp. 349-352.

[BL04] Bowers S. and B. Ludäscher, 2004. An ontology-driven framework for data transformation in scientific workflows, In Proc. of Workshop on Data Integration in the Life Sciences (DILS), LNCS, vol. 2994, pp. 1-16.

[BTWL04] Bowers S., D. Thau, R. Williams, and B. Ludascher, 2004. Data procurement for enabling scientific workflows: On exploring inter-ant parasitism, In Proc. of Workshop on Semantic Web and Databases (SWDB), LNCS, vol. 3372.

[DC] Dublin Core Metadata Initiative. http://dublincore.org.

[EML] Ecological Metadata Language. http://knb.ecoinformatics.org/software/eml/.

[EOL] Encyclopedia of Life. http://eol.sdsc.edu/.

[FGDC] Federal Geospatial Data Committee. http://www.fgdc.gov.

[SP99] Stockwell D.R.B. and D. Peters 1999. The GARP Modeling System: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science 13 (2): 143-158.

[GEON] The Geosciences Network. http://www.geongrid.org

[JBBS01] Jones, M.B., C. Berkley, J. Bojilova, and M. Schildhauer, 2001. Managing Scientific Metadata, IEEE Internet Computing 5(5): 59-68.

[KEPLER] Kepler: An Extensible System for Scientific Workflows, http://kepler.ecoinformatics.org

[KNB] Knowledge Network for Biocomplexity. http://knb.ecoinformatics.org.

[LL00] E. A. Lee and S. Neuendorffer. 2000. "MoML - A Modeling Markup Language in XML, Version 0.4," Technical Memorandum UCB/ERL M00/12, University of California, Berkeley, CA 94720, March 14, 2000. http://ptolemy.eecs.berkeley.edu/publications/papers/00/moml/.

[OWL] Web Ontology Language. http://www.w3.org/TR/owl-features/.

[PTOLEMY] Ptolemy II, http://ptolemy.eecs.berkeley.edu/ptolemyII/.

[ROADNET] Real-time Observatories, Applications and Data Management Network. http://roadnet.ucsd.edu.

[SDM] Scientific Data Management Center. http://sdm.lbl.gov/sdmcenter/.

[SEEK] Science Environment for Ecological Knowledge. http://seek.ecoinformatics.org.

[WSDL] Web Services Description Language (WSDL) 1.1, W3C Note 15 March 2001, http://www.w3.org/TR/wsdl.

LTER

Network
Information
System

LTER Network Information System Strategic Plan

Network Information System Advisory Committee – Release Candidate Version 2.8

Henshaw, Donald - AND (chair)
Benson, Barbara – NTL
Boose, Emery - HFR
Brunt, James - LNO
Gage, Stuart - KBS
Harmon, Mark - AND
Kratz, Timothy - NTL
McCartney, Peter - CAP
Michener, William - LNO
Peters, Debra - JRN
Ross, Robin – PAL
Servilla, Mark - LNO
Vande Castle, John - LNO
Waide, Robert - LNO

Comments to:
nisadvise@LTERnet.edu
Re: Release Candidate Version 2.8
15 July 2005

1 Introduction

Data and synthesis products resulting from research activities across the Network of LTER sites and their partners are raw material for the generation of new ecological knowledge. The LTER Network Information System (NIS) seeks to accelerate the development and use of these products. Functionally, this will be achieved by implementing shared standards, confederating site data repositories, developing software tools, training, and support to integrate and provide compatibility across the Network of sites, institutions, and researchers. The NIS will serve the LTER scientific community and collaborators, and provide a "portal" to LTER data products for the broader scientific community, natural resource managers, policymakers, and the general public. Conceptually, the NIS is a process that will enhance flows of information among data, synthesis and knowledge about ecological systems, will depend on the science and Information Management (IM) participants at the site and network levels for implementation, and will both rely upon and contribute to informatics expertise and IM systems outside of LTER (Figure 1).

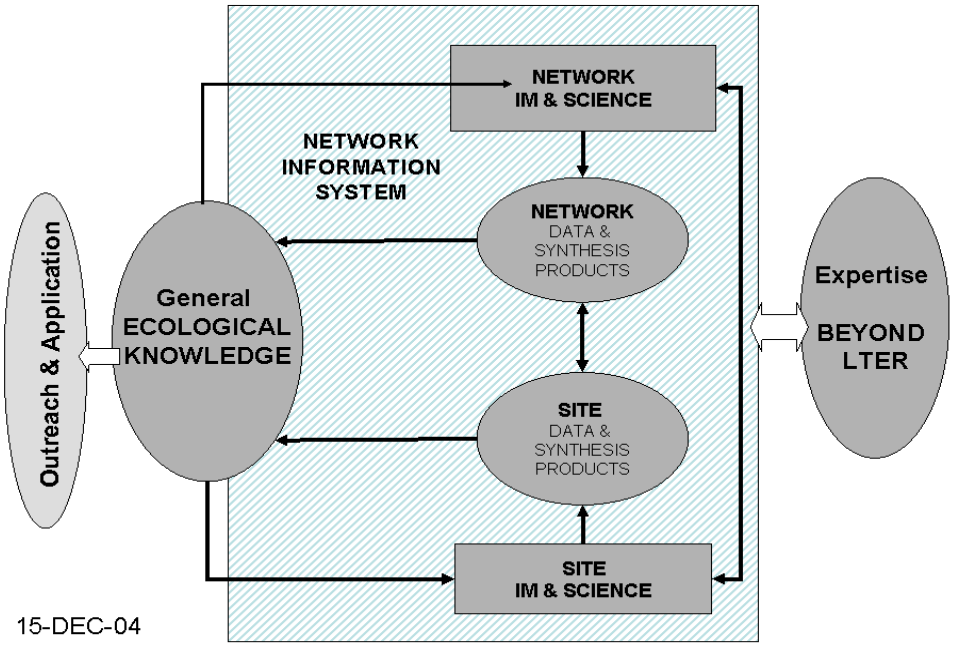


Figure 1 – Schematic diagram of NIS conceptual framework.

In this process, the NIS will seek to balance development of short-term, priority products with development of long-term, generic solutions and improved network-wide infrastructure. With these functional and conceptual views in mind, this plan defines the mission and goals of the NIS effort and elaborates the strategies necessary to make progress over the next 4 years.

2 Mission Statement

The mission for the LTER Network Information System (NIS) is to provide the Information Management and Information Technology infrastructure to facilitate and promote advances in collaborative and synthetic ecological science at multiple temporal and spatial scales.

3 Goals

To realize the stated mission it is recognized that the Network Information System (NIS) needs to improve the overall quality of site information systems, confederate these systems, and then enable the use of these systems for the advancement of scientific knowledge. In this highly linear representation of NIS goals, the tasks at hand are seen as completely interdependent – i.e., knowledge cannot be achieved without confederated databases. Realizing that progress must be made simultaneously on many fronts, a non-linear planning approach to produce measurable progress along a tiered trajectory has led to the following goals that support the core research objectives of the LTER Network as well as being vital to fulfilling the stated NIS mission. These goals are structured along the natural progression from data to knowledge as described above. Strategies that support these goals have been elaborated (Section 5) and focus on promoting standards and providing software tools, training, and support.

3.1 Data: Increase data quality through standard approaches

Data are the legacy of the LTER Network. NIS will seek to improve the overall quality of site infrastructure to efficiently manage, archive and curate this heritage in a standardized way. NIS will need to promote and support standardization in the management of information content including primary data observations, models (conceptual design, implementation code, input and output data), secondary data products, syntheses of existing data, and knowledge representation (ontologies).To build on the strength of the existing autonomous but heterogeneous site data repositories the NIS will need to facilitate improvements in metadata quality and site information systems in a tiered and measurable approach.

3.2 Synthesis: Increase data available for synthetic activities

Discovery, access, and use of LTER data will be enabled by facilitating the communication and interoperability between heterogeneous systems. NIS will need to develop and deploy applications that accommodate LTER information content, including an on-line data catalog and applications to exploit discovery of these data for reuse. NIS will need to develop and maintain shared software solutions for the integration of local site information systems(middleware). NIS will need to support multiple interfaces to 1) search and discover all site databases and 2) make the information resources of the entire Network available to the tools used by scientists in synthetic research.

3.3 Knowledge: Increase knowledge discovery through synthesis

NIS will be used to make new scientific discoveries and provide new data resources by enabling semi-automated integration, modelling and forecasting. NIS will need to support the creation of Network-based synthetic information products by implementation of relational database technology, shared middleware, community-based applications and scientific collaboration. NIS will need to participate in the development and/or adoption of standards for managing metadata associated with models and forecasting. To make an impact at this level, NIS will need to pursue this goal in close coordination/collaboration with the broader IT/IM and ecoinformatics communities.

4 Implementation

4.1 Process

The LTER Network Coordinating Committee (CC) established the Network Information System Advisory Committee (NISAC) in April 2003 to formalize the process of developing a Network Information System (NIS) for LTER. This committee, made up of representatives from site principal investigators (PI), information managers (IM), and the LTER Network Office (LNO), took on as its first responsibility the development of a strategic plan for the NIS. Sites, LNO, the CC, the IMC, and the NISAC each has a role in the successful design, development, and implementation of NIS. As part of this process NISAC has defined these roles and responsibilities.

4.2 Roles

NISAC – NISAC is responsible for planning and evaluation of NIS activities. NISAC will present clear and effective goals and strategies to the CC and Network community to educate and garner support for NIS initiatives. The NISAC will present NIS advances to funding agencies for continued support of activities and define roles and create partnerships. Additionally, NISAC will regularly assess LNO/SITE performance and progress towards NIS goals and evaluate the effectiveness of strategies toward achieving NIS goals. NISAC will modify the strategic plan and take iterative action as necessary.

LNO – LNO will lead the design, development, deployment, and support of NIS infrastructure. LNO will provide timely information and demonstration of deliverable products to the Network and broader communities. LNO will collaborate with sites to develop NIS data products, new resources for implementation of Network standards, and Network specific information technology and training. LNO will demonstrate NIS capabilities to scientists, information managers, and other stake-holders to gain broader participation in Network data synthesis.

CC – The coordinating committee will direct NISAC to undertake activities and approve and endorse the goals and strategies elaborated in the NIS strategic plan.

SITE – Sites will implement Network standards, provide access and timely updates to sites research data and metadata in formats that have been agreed upon by the CC. Sites manage the sites research data, including assuring data quality, preparing metadata for sites research data, maintaining a local archive of sites research data, and providing for security of site research data. Sites participate in NIS as a node in the distributed data network, being both data consumer and data provider.

IMC – The Information Management Committee will review NIS strategic plan action items directly affecting the sites and communicate comments to NISAC. The IMC will plan for the implementation of Network standards at sites including the development of cost and time estimates. The **IMEXEC** will be responsible for survey, accumulate, and communicate status of progression of sites through the tiered trajectory to NISAC.

5 Strategies

Because of the heterogeneity across LTER site information systems, and the need to balance short-term and long-term products, the approaches engaged to achieve the stated goals must necessarily be diverse. For example, in support of one such scientific data product, an approach of harvesting data and aggregating in a central location might make the most of available resources while the nature of a second scientific data product might well lend itself to a more distributed query (Figure 2).

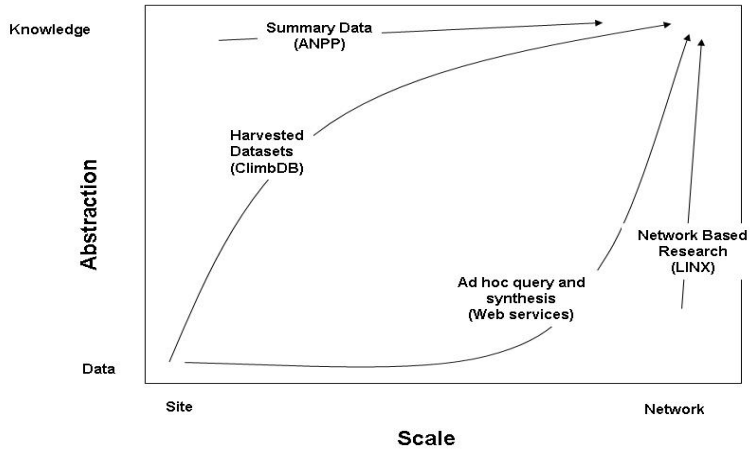


Figure 2 – Graphical representation of the diversity of approaches to synthesis.

The proposed strategies that follow are aligned along a trajectory from site needs to network solutions and follow the natural progression from data to knowledge corresponding to the 3 goals of the NIS (Figure 3). The primary focus in each of these 3 task areas falls on the information managers, the developers, and the research scientists respectively for the data, synthesis, and knowledge goals.

5.1 Data: Increasing data quality through standard approaches

NIS will support standardization in the development and management of information content at the sites through guidance, resources, training, and support.

Strategies**:**

5.1.1 Increasing the quality of data and metadata by adopting, developing, and implementing standards.

5.1.2 Increasing support for local site information management and information technology to accommodate new standards, infrastructure, etc.

5.1.3 Providing training, guidance, and workshops on new information technology and information management to sites.

5.2 Synthesis: Increasing data available for synthetic activities

NIS will develop and deploy applications that accommodate LTER information content, including an on-line data catalog and applications to exploit these data for discovery of information.

Strategies:

1. Increasing the use of available data by developing and deploying basic applications and interfaces for discovering and accessing LTER data.
2. Overcoming site heterogeneity issues by developing, adopting, or adapting middleware to enable efficient access to site data.
3. Supporting workshops for network developers to integrate specific information technologies.

5.3 Knowledge:Increasing knowledge discovery through synthesis

NIS will support the creation of Network-based synthetic information products through the use of relational database technology, shared middleware, community-based applications and scientific collaboration.

Strategies:

1. Providing education and training to researchers on NIS applications.
2. Supporting exploitation of LTER information products through development of advanced applications.
3. Supporting the creation of Network-based synthetic information products through application development, scientific collaboration, and product-oriented workshops.

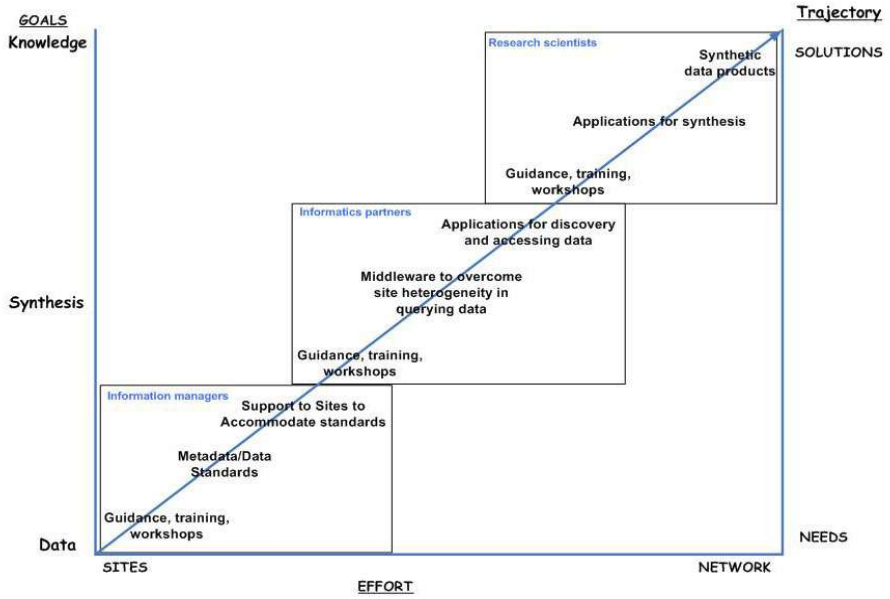


Figure 3: NIS strategies aligned along a trajectory from site needs to network solutions.

Review Criteria for LTER Information Management Systems

Version 1.0 12 April 2005

INTRODUCTION

This document was created by the LTER Information Managers Committee, vetted by the LTER Network Information System Advisory Committee, and approved by the LTER Coordinating Committee on 6 April 2005. The document is intended to serve as a reference for formal reviews of LTER sites as well as for informal self-assessment and planning.

The Information Management System at an LTER site encompasses hardware, software, and people to store and deliver scientific information (data and metadata, where data may include both tabular and spatial representations). The goal of an Information Management System is to support site and network science by (1) facilitating access to data and metadata by LTER scientists, the scientific community, and the public, and (2) by ensuring the integrity, security, and usability of those data and metadata for future generations.

The review criteria below focus on functionality rather than specific implementation. A successful Information Management System is created and maintained through the coordinated efforts of the Information Manager, other information technology staff, field and laboratory technicians, researchers, and site management. The Information Management System should be evaluated as an integral part of the overall LTER program at a site.

LTER sites are expected to meet criteria designated as *shall*, and to either meet or show measurable progress toward criteria designated as *should*. Reviewers should realize that the bar has been set high in this document to encourage excellence. Sites are not expected to score perfectly in all areas but should demonstrate steady progress toward network goals as outlined and prioritized below.

REVIEW CRITERIA

A. Information Management System design and implementation

1. Scope
    a. Data and metadata shall be made available online as specified and prioritized in the LTER Network Data Access Policy [1].
    b. Sites shall have a procedure for making data stored offline (e.g., large GIS, remote sensing, or modeling datasets) accessible to the scientific community.
    c. The Information Management System shall include an up-to-date list of publications supported by the site LTER program.

    d. Inclusion of catalogs of non-electronic materials managed in support of LTER research (samples, specimens, documents, photographs, etc) is encouraged.
2. Design
    a. The Information Management System should conform to current best practices for critical design features such as data and metadata encoding, short-term backup, long-term media and format migration, system administration, security, scalability, and query capability.
    b. Site data and metadata shall be backed up regularly and copies stored offsite to protect against disaster.
    c. Sensitive data (such as personal information or location of endangered species) shall be protected against misappropriation and misuse.
    d. Innovations in design or methods, especially where suitable for use by other sites, are encouraged.
3. Web page
    a. Data, metadata, and publication list shall be well organized, readily located, and easily accessed from the site web page.
    b. Site web page should conform to LTER Network recommendations [2].
    c. Innovations in web page design and Information Management System interface, especially where suitable for use by other sites, are encouraged.
4. Documentation
    a. Information Management System architecture, procedures, and protocols shall be clearly documented and documentation shall be sufficient to maintain continuity if there is a turnover of personnel.
    b. The Information Management System shall include an up-to-date list of current and completed LTER-related research projects.
    c. Site shall have a management plan for the Information Management System indicating how critical tasks are accomplished by site personnel.
5. Review
    a. Site management shall conduct an annual internal review of the site Information Management System.
    b. Site shall demonstrate measurable progress toward addressing recommendations from previous internal and external reviews.

B. Information Management System support for site, network, and community science

1. Integration with site science
    a. All stages of project development from initial project design to final archiving of data and metadata should be integrated into the Information Management System.
    b. Researchers should be able to make effective use of the site Information Management System.
2. Policies
    a. Site data release, access, and use policies shall comply with LTER Network policies [1].
    b. Site policies shall be clearly stated on the site web page.

3. Metadata
    a. Metadata shall be of sufficient quality and completeness to ensure long-term (> 20 years) usability of data [3].
    b. Metadata shall be EML-compliant at level 2 (discovery) [4]. Metadata should be EML-compliant at level 5 (integration) [4]. Site EML shall comply with LTER best practices [4].
4. Data
    a. Data integrity shall be protected by appropriate quality control procedures.
    b. Data access should be tracked in accordance with the LTER Network Data Access Policy [1].
5. Contribution to LTER Network and community activities
    a. Site shall have consistent representation at the annual LTER Information Managers Committee meeting.
    b. Site shall contribute relevant data and metadata to Network Information System modules approved by the LTER Coordinating Committee (ClimDB, SiteDB, etc).
    c. Participation by the Information Manager in other LTER activities such as committees, workshops, and tool development; in community activities such as review teams, panels, and collaborations with informatics partners; and in related research activities such as developing proposals and publications, is encouraged.

REFERENCES

[1] LTER Network Data Access Policy & LTER Network Data Use Agreement.
See LTER Intranet (http://intranet.lternet.edu) for current versions.

[2] Recommendations of the Information Managers Committee will be available on the LTER Intranet (http://intranet.lternet.edu) by fall 2005.

[3] Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications **7**:330-342.

[4] EML Best Practices for LTER Sites.
See LTER Intranet (http://intranet.lternet.edu) for current version.

Data Access Policy Review

- Charged by CC in Fall 04 meeting to update DAP to be consistent with other NSF programs and develop appropriate use guidelines
- Review completed December 2004.
  - Summary of findings
  - Appendix 1: revised Data Access Policy V 3
  - Appendix 2: General User Agreement

Data Release

- Data from publicly funded research in the U.S. LTER Network, totally or partially from LTER funds from NSF, Institutional Cost-Share, or Partner Agency or Institution where a formal memorandum of understanding with LTER has been established is to be released either:
  - Publicly available (within 2 years, most data)
  - Restricted users (sensitive data, human subjects act, prior licensing restrictions – requires justification)
- All data are released with metadata compliant to LTER standards – identification, citation, access, use
- All metadata are public.

What does this cover?

- In spirit, ALL DATA
- In practice, priorities may be defined:
  1. Primary observations or secondary (value added products) from core research activities with direct or leveraged LTER funds
  2. Acquired (tertiary) data, student data, SLTER data, ‘legacy’ data
  3. Raw, un processed data; interim statistical, GIS, modeling outputs.

Registration

- Information for tracking data use and notification among relevant parties
  - Contact information
  - Acceptance of the Data User Agreement
  - Statement of intended use
- Most of this information will be collected through online registration

General Use Agreement

- Standard template for specifying use guidelines for public LTER data
- Derived from open content forums
- Very generic language to enable use beyond LTER
- References parties identified in the accompanying metadata:
  - Creater, Owner, User, Contact, Provider

Terms

- Acceptable use
- Redistribution
- Citation
- Acknowledgement
- Notification
- Collaboration

Some Caveats

- Facts cannot be copyrighted, only creative products. Data fall in the boundary.
- Products of federal agencies cannot be copyrighted.
- Use guidelines are considered ethically, probably not legally, binding.

The Canopy Database Project

Tools for Research & Information Integration

http://canopy.evergreen.edu

Judy Cushing, Nalini Nadkarni
Mike Finch, Anne Fiala, J. Lee Zeman and others
The Evergreen State College

Collaborating Ecologists	Collaborating Computer Scientists	Collaborating LTER Information Mangers
Van Pelt, Bond, Dial, Ishii, Keim, Parker, Shaw, Sillett, Sumida, et al	Dave Maier, Lois Delcambre, Travis Brooks (PSU)	Nicole, Kristin, Ken, Eda, Jonathan, James Brunt and others?

NSF CISE and BIO 04-0417311, 03-019309, 01-31952, 01-9309 99-75510, 9630316, 93-07771

Canopy DB Vision

Templates -- *Domain Specific Data Structures*

Database & Component Technology can help:

1. Generate, maintain metadata,
2. Validate, archive data,
3. Increase researcher productivity,
4. Facilitate ecological synthesis.

Scientific Case Study –

Thousand Year Chronosequence (1kcs)

1. Within-tree & -stand structure (across age classes)
2. Structure-function Synthesis

Collaboration with LTER IMs & Sites

1. Grasslands Data Integration JRN, SEV, SVS
2. Morpho PlugIns – B. Bond (PI @ HJA PI)
3. Forest Canopy Visualization – Luq
4. Eco-Informatics for Resource Managers
5. SSDBM Panel on CS-IM Collaboration (Kristin, Peter, Matt, Amarnath, me)
6. OSU Eco-Informatics NSF-IGERT
7. Managing Field Databases (in Access)….

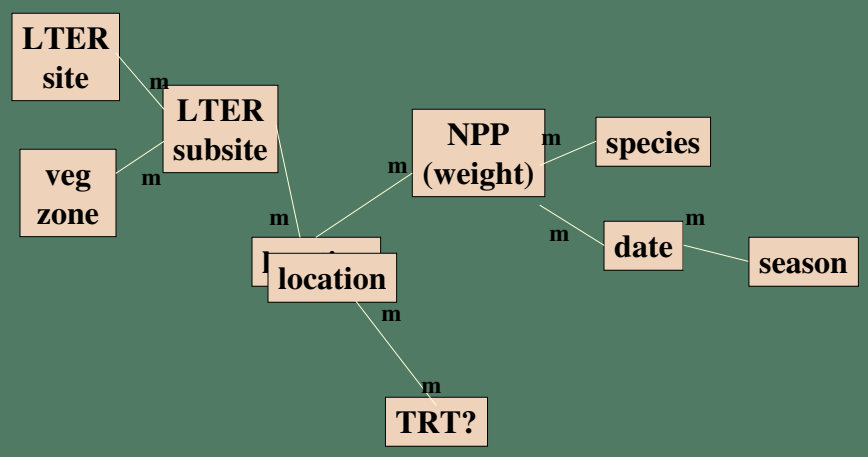
Grasslands Biomass Data Integration Case Study

Scientific outcome -- Cross site analysis of NPP (JRN, SEV, SGS)

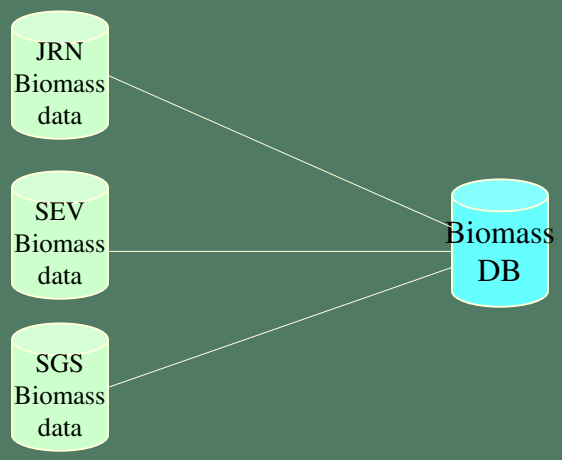
1. **Compare production & species richness, want g/m2 per species per quadrat & number of species per quadrat.**
2. **Compare biomass over biomes - sum g/m2 over biomes at each site.**

Eco-informatics/CS outcome – published ecology data integration case study

Grasslands Biomass Data Integration Schema



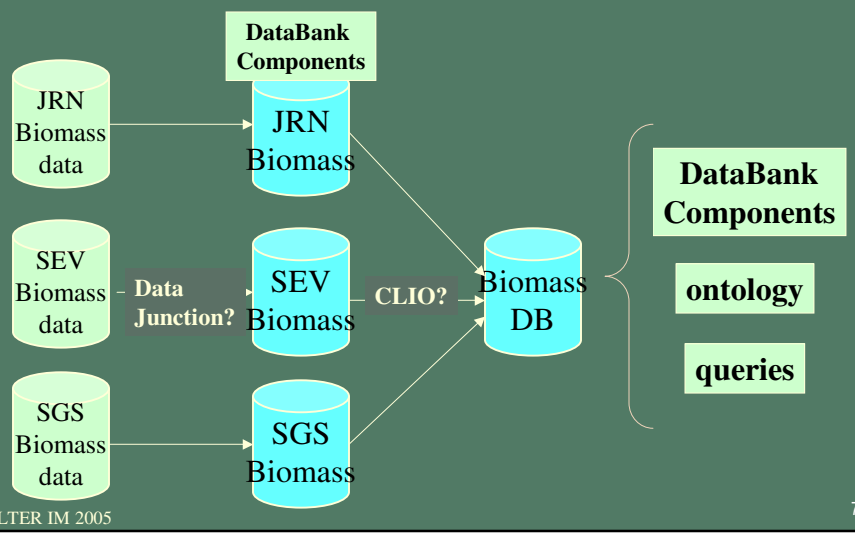
Grasslands Biomass Data Integration Experiment



By hand data conversion is 'hard',
Includes non-DB calculations

Grasslands Biomass Data Integration Strategy

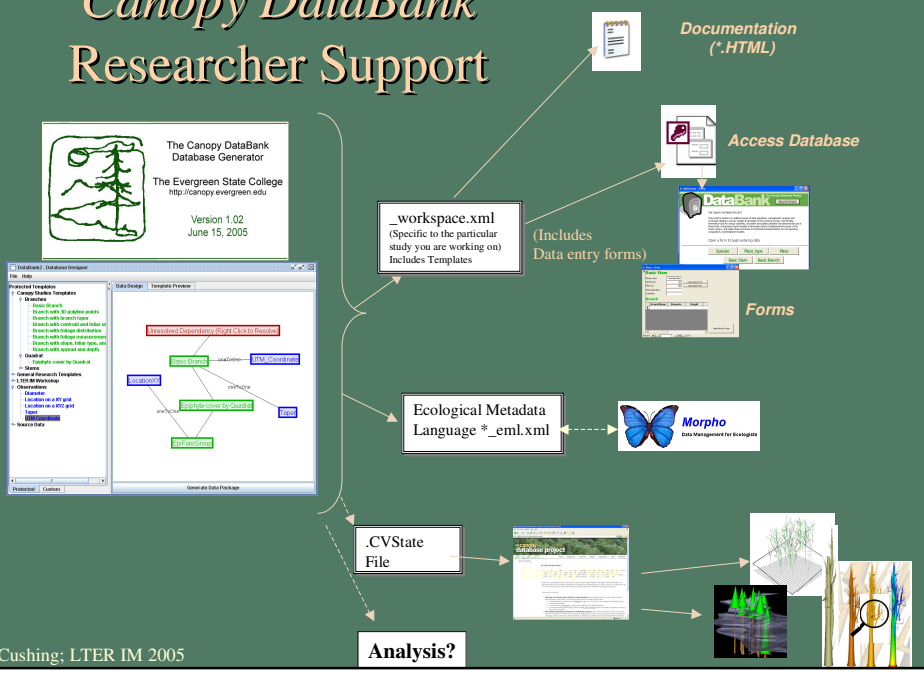
Staged data conversion easier?



Cushing; LTER IM 2005

7

Canopy DataBank Researcher Support

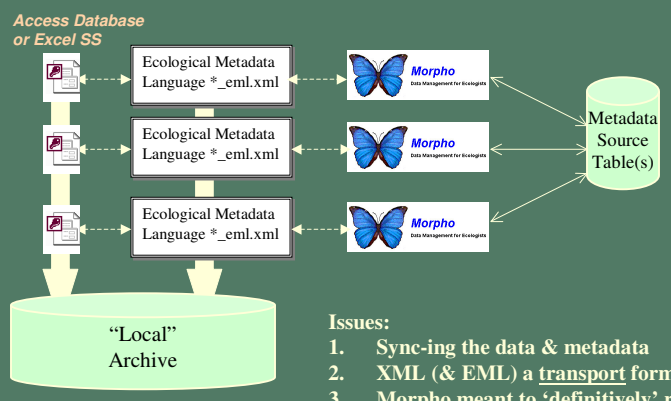


Cushing; LTER IM 2005

8

Building a Structural & Semantic Metadata Record of the Research Study

metadata from the data source and data use iteratively,
& re-use & share metadata "source"



Issues:

1. **Sync-ing the data & metadata**
2. **XML (& EML) a transport format**
3. **Morpho meant to 'definitively' record**

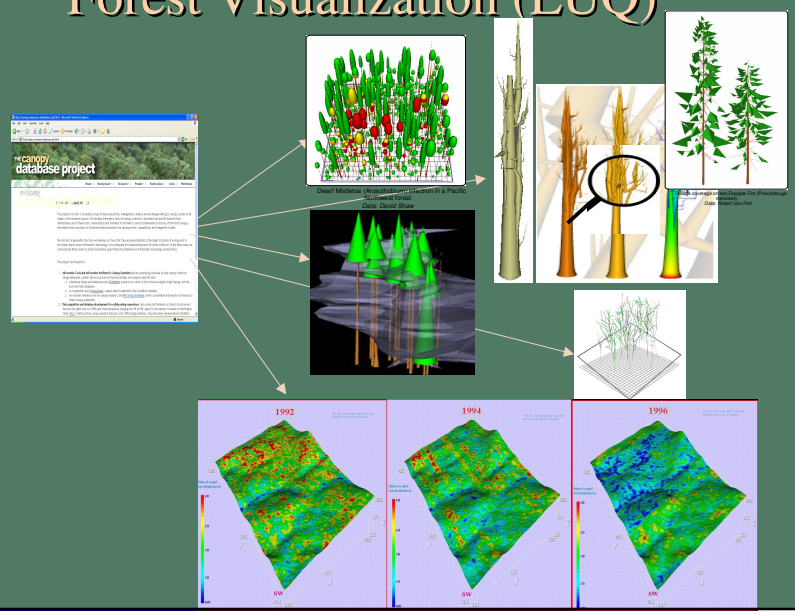
Morpho PlugIns

some w/ B. Bond Lab (an HJA PI)

1. Metadata Export to Morpho (from DataBank, Access, Excel)
2. Source/Lookup Table for Morpho
3. Flexible profile for Morpho directory (allows source table sharing) *
4. Data package copy of Morpho packages *

***not strictly speaking plug-in**

Forest Visualization (LUQ)



11

Eco-Informatics for Decision Makers Advancing a Research Agenda

NSF IIS 0505790 (DG)

www.evergreen.edu/bdei

December 2004 Workshop
The Evergreen State College

Judy Cushing
The Evergreen State College
judyc@evergreen.edu

Tyrone Wilson
USGS/NBII
tyrone_wilson@usgs.gov

Alan Borning (UWa), Lois Delcambre (PDX),
Geoff Bowker (Santa Clara), Mike Frame (USGS/NBII),
John L. Schnase (NASA), William Sonntag (EPA),
Janos Fulop (Hung. Science Acad.), Carol Hert (UWa),
Eduard Hovy (USC), Julia Jones (OrSU), Eric Landis,
Charles Schweik (UMass-Amherst),
Larry Brandt, Valerie Gregg, Sylvia Spengler (NSF)

17 co-Authors,
27 Speakers,
47 participants

Cushing; LTER IM 2005

12

Grasslands IMs and Eda
Ken, Kristin, Nicole

Meet here:

Sunday 4-5:30 (then dinner!)

Meet Sherbrook 2?:

Monday 8:30-5pm

Long Term Ecological Research
Network Information System

LTER Metacat
Advanced Query Interface

LTER Information Manager's Meeting
Montreal, Canada
4-7 August 2005

Mark Servilla (LNO)

Goals & Products

- Initial design - 2004 IM Meeting, Portland
- Supported under the LTER NISAC
- More expressive queries to Metacat
- Interface provides
  - Simple query form
  - Advanced query form
    - Subject
    - Author
    - Spatial
    - Taxonomic
  - Simple “browse catalog” (KNB like)

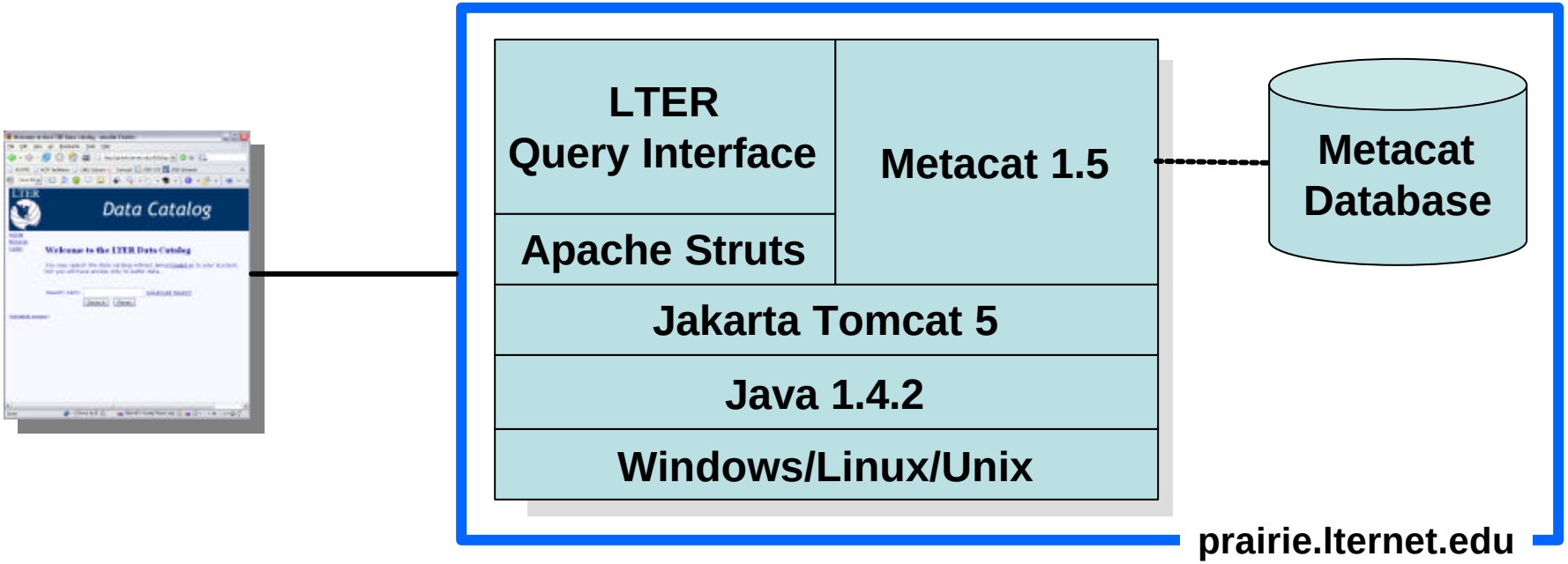
Goals & Products cont.

- Decouple from Metacat to enhance portability
  - Runs as a separate web application (as a servlet in a Tomcat container)
- Utilize standard patterns and designs
  - Apache Struts MVC design
- Parameterized for site-specific skins (in progress)

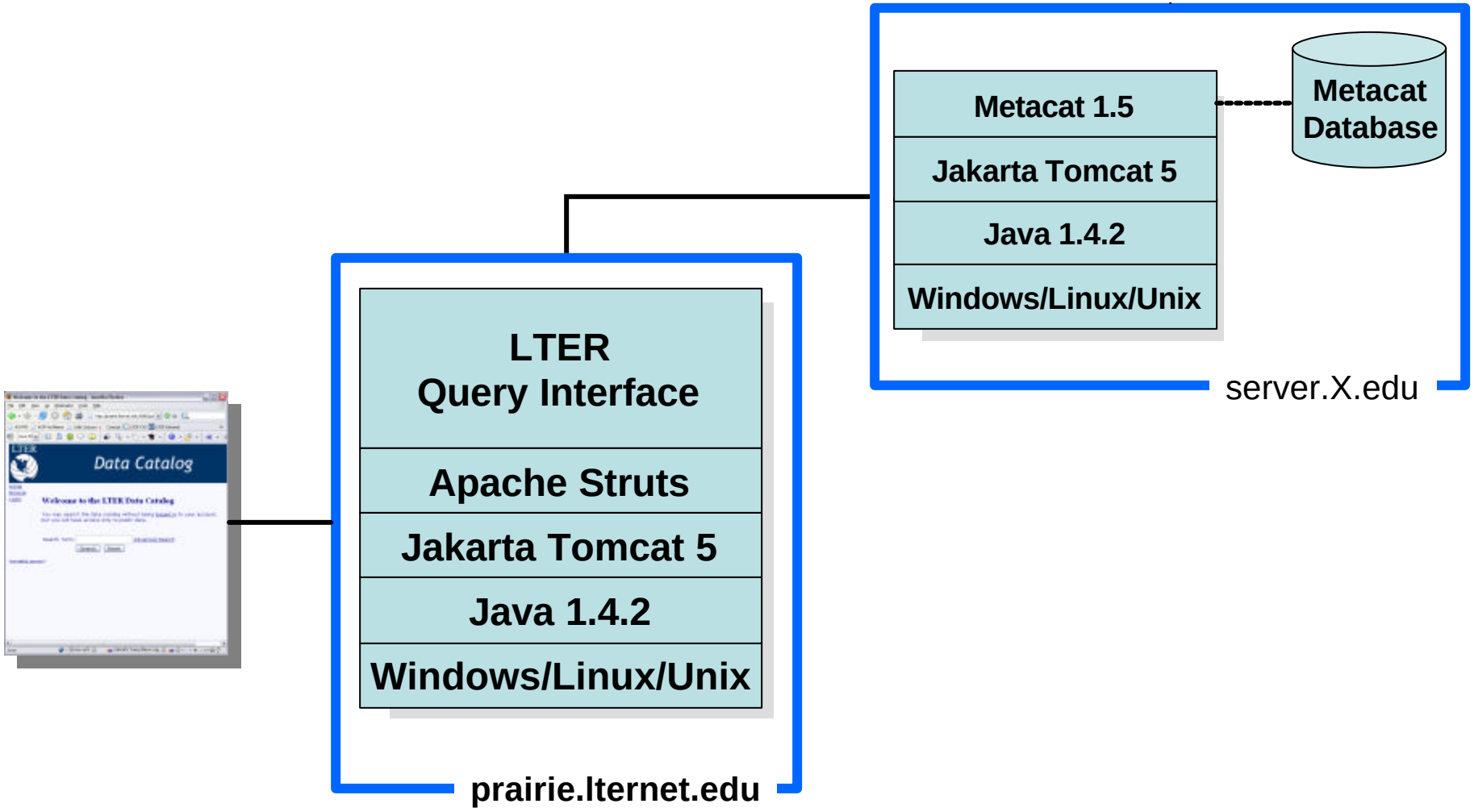
Metacat Notes

- Performance issues are being addressed by metacat-dev
  - Separate active versus deprecated xml_nodes
  - Generate database indices for key columns
  - DBA configurable indices
  - Enhance xml_node query syntax
  - Improve host hardware configuration/recommendations (e.g., up memory, clustering, etc…)

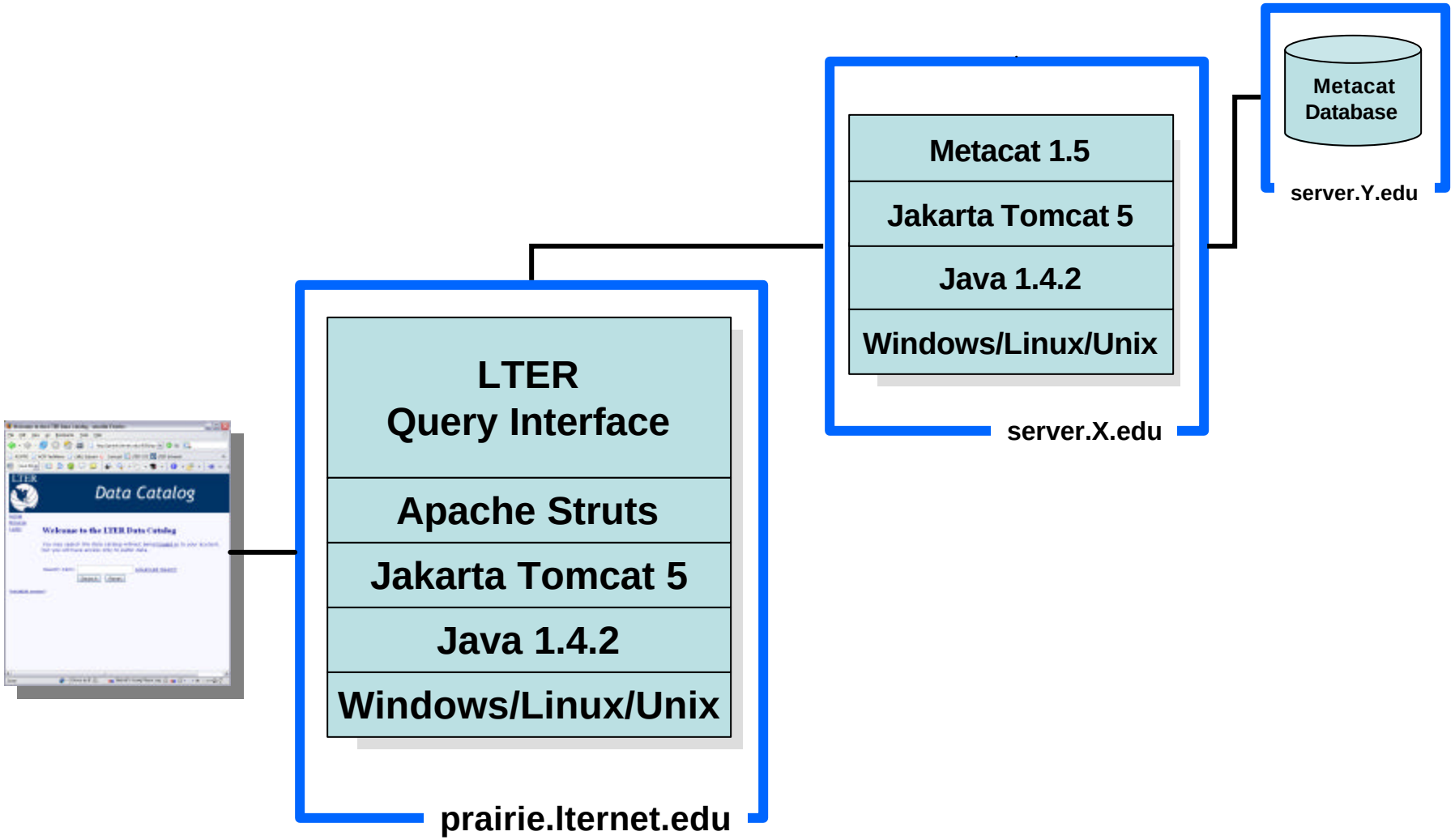
Architecture



Architecture "X"



Architecture "X-Y"



LTER Custom Unit Registry

Demonstration

Long Term Ecological Research
Network Information System

LTER Custom Unit Registry Project

LTER Information Manager's Meeting
Montreal, Canada
4-7 August 2005

Mark Servilla (LNO) & Karen Baker (PAL)

Goals & Products

- Custom Unit Registry
  - Completed
    - Extensible “unit type” and “unit” RDBMS schema
    - Web-based searchable registry
    - Tabular and EML/XML output
    - Scope-based and hierarchical use designation
  - Planned
    - Entry form for submitting new “unit types” and “units”
    - Entry form for editing “unit types” and “units”
    - Web-service connector
    - User Authentication/Authorization

Goals & Products cont.

- Community Process
  - Facilitate the definition and exchange of "measurement unit" information
  - Supports the collections, exchange, and adoption of "units" through a formal acceptance process
  - Ensure compatibility with community standards
  - Develop communication pathways for input from domain experts
  - Establish "Custom Unit Best Practices"

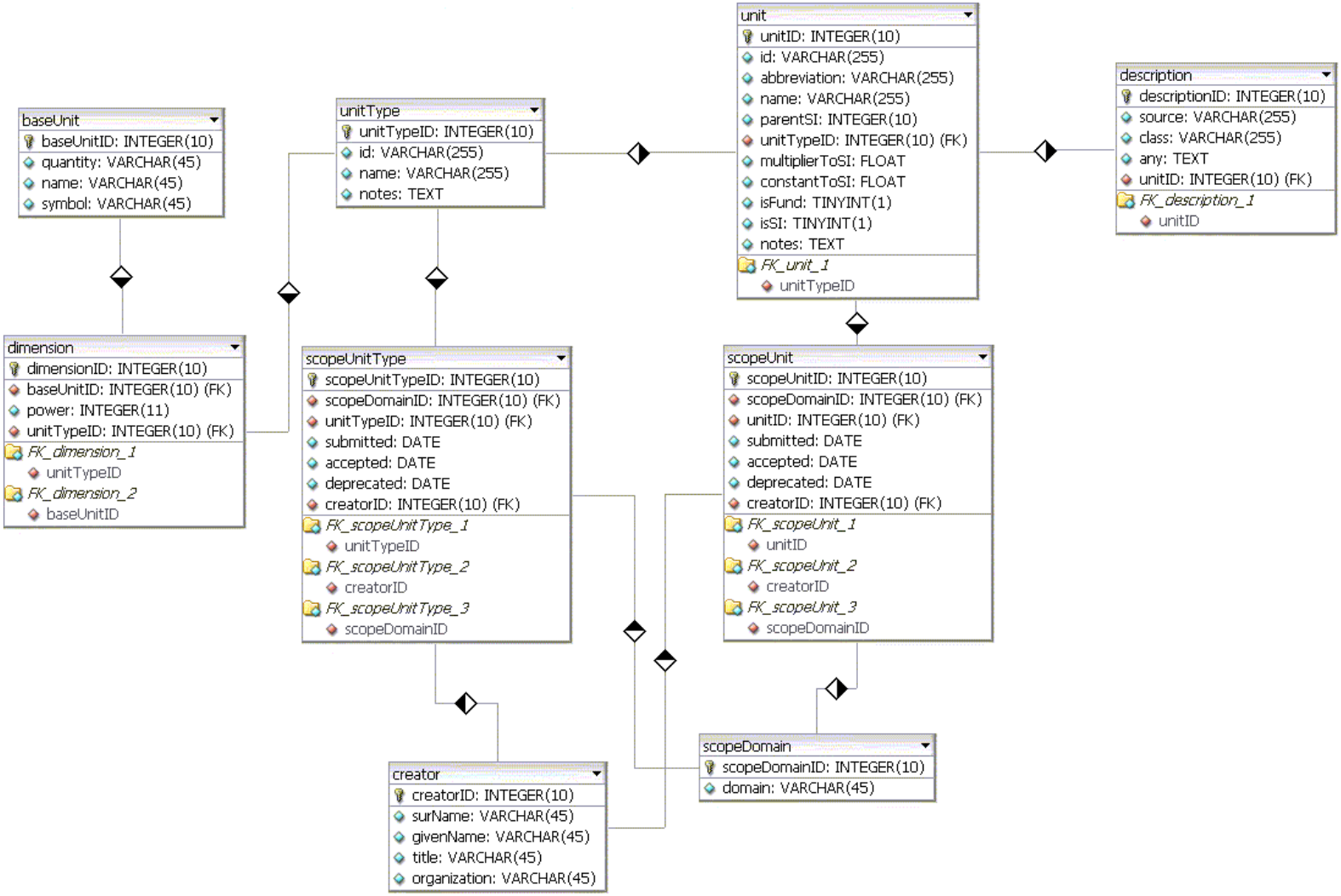
Schedule

- May - July 2005
  - Functional database schema populated with
    - EML-2.0.1 unit type and unit definitions
    - Custom units from FCE, GCE, LNO, PAL, and SBC
  - Web-based searchable interface
    - http://fire.lternet.edu/customUnit
    - Unit type and unit keyword searching
    - EML compatible XML output
- August - December 2005 (to be decided)
  - Entry and editing forms
  - Web-services connector
  - Security

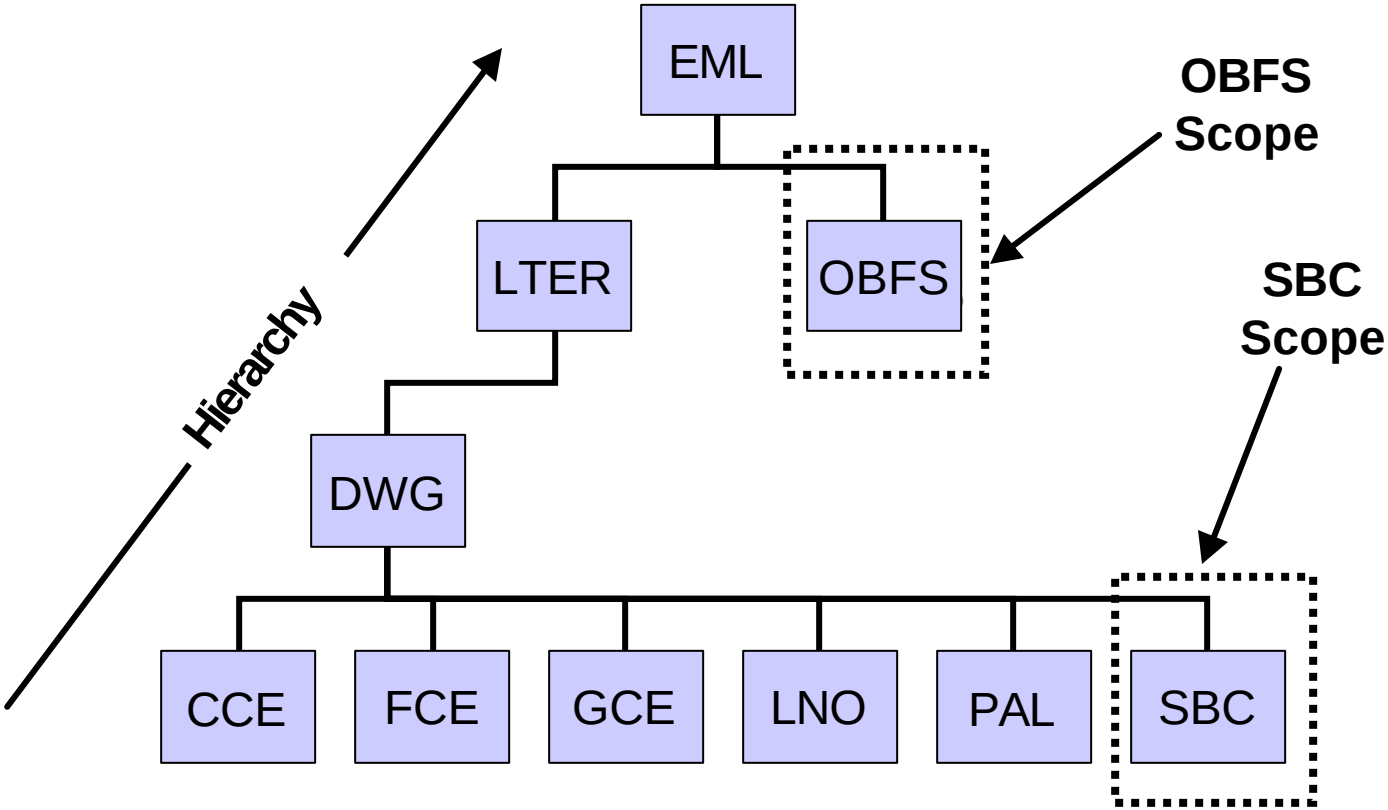
Design Team

- CCE – Lynn Yarmey
- FCE – Linda Powell
- GCE – Wade Sheldon
- LNO – Mark Servilla and Inigo San Gil
- PAL – Karen Baker, Shaun Haber, and Florence Millerand
- SBC – Margaret O'Brien

Database Schema



Unit Scope and Hierarchy



LTER Custom Unit Registry

Demonstration

Long Term Ecological Research
Network Information System

<unitList>…</unitList>

LTER Information Manager's Meeting
Montreal, Canada
4-7 August 2005

Mark Servilla (LNO)

System Internationale

- Base Units - Quantity versus Unit
  - Quantity is the amount that is being measured
  - Unit is a determinate quantity adopted as a standard of measurement

	SI base unit	
Base quantity	***Name***	***Symbol***
length	meter	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	kelvin	K
amount of substance	mole	mol
luminous intensity	candela	cd

SI cont.

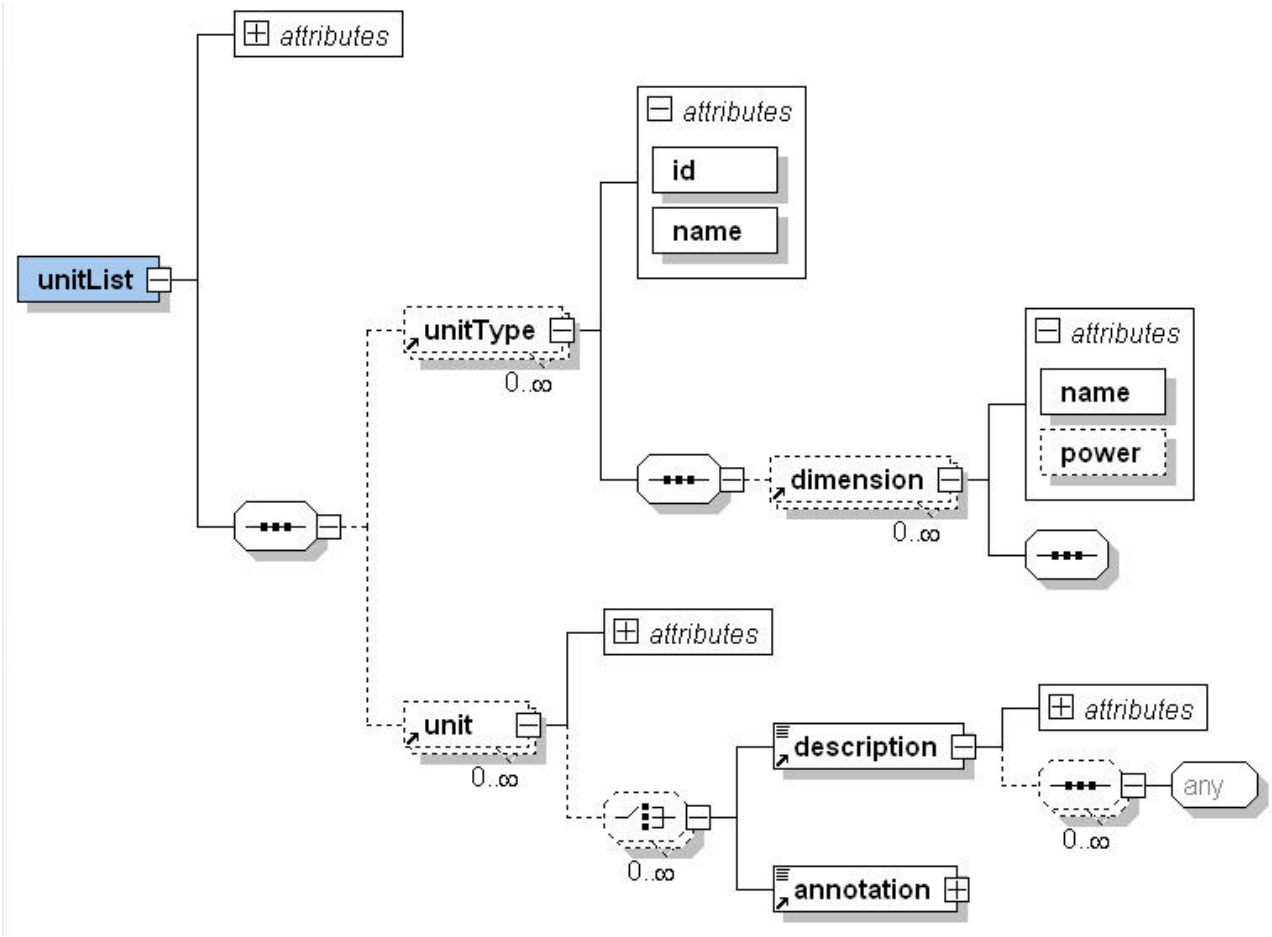
- Derived Units
  - "formed as products of powers of the base units according to the algebraic relations linking the quantities concerned"
- Examples
  - Speed :: metersPerSecond – m/s
  - Acceleration :: metersPerSecondSquared – m/s^2
  - Volumetric Rate :: cubicMetersPerSecond – m^3/s

STMML

- Scientific, Technical, Medical Markup Language
  - <unitList>
  - <unitType>
  - <unit>
- Example

```
<unitList>
     <unitType id="volumetricRate" name="volumetricRate">
        <!-- EML-2.0.1 -->
        <dimension name="meter" power="3"/>
        <dimension name="second" power="-1"/>
     </unitType>
     <unit id="cubicMetersPerSecond" abbreviation="m³/s"
        name="cubicMetersPerSecond" unitType="volumetricRate">
        <!-- LNO-LTER -->
        <description>
               cubic meters per second
        </description>
     </unit>
</unitList>
```

STMML Schema



STMML Rules

- There exist 7 base SI units
- Derived SI units can be formed by any combination of products of powers of the base units
- SI units cannot have a "parent SI" unit
- SI units cannot have a "multiplier" or "constant" to derive another SI unit
- Non-SI units must have a relationship to an SI unit, and this SI unit is defined as the "parent SI" unit
- Non-SI units must have a scalar quantity, the "multiplier", that can be used to convert the non-SI unit quantity to the SI unit quantity
- Non-SI units may have a scalar quantity, the "constant", that can be added to the non-SI unit (in combination with the "multiplier") to convert the non-SI unit quantity to the SI unit quantity

Beyond the eml-unitDictionary

- Current limitations
  - Error correction
  - Extensibility
- Future directions
  - dictionaries and registries
    - LTER Custom Unit Registry Project
  - LSID (Life Sciences IDentifier)
    - urn:lsid:lternet.edu:u432:2
  - Ontologies (as related to Attributes)

Long Term Ecological Research
Network Information System

LTER Grid Pilot Study

LTER Information Manager's Meeting
Montreal, Canada
4-7 August 2005

Mark Servilla (LNO)

Pilot Study Goals

- Demonstrate the vision of the LTER Grid to the LTER community
- Demonstrate how grid and middleware technologies can be applied to enhance collaboration and sharing within and external to the LTER community
- Demonstrate an effective working relationship with the LTER Network office, NCSA, ISI, and GRIDS Center staff
- Demonstrate the potential of super computers for LTER science
- Collect requirements for a detailed design of the LTER Grid

Strategies

- Contrived scientific experiment
- Short duration (6-month) end milestone
- Demonstration of final product
- Directly involve stake-holders
- Identify key and usable Grid infrastructure

LTER Grid Pilot Study

- LTER Network Office
  - University of New Mexico
- NCSA
  - University of Illinois at Urbana-Champaign
- ISI
  - University of Southern California
- KBS LTER (Dr. Stuart Gage)
  - Michigan State University
- SEEK
  - NCEAS at the University of California, Santa Barbara

LTER Requirements

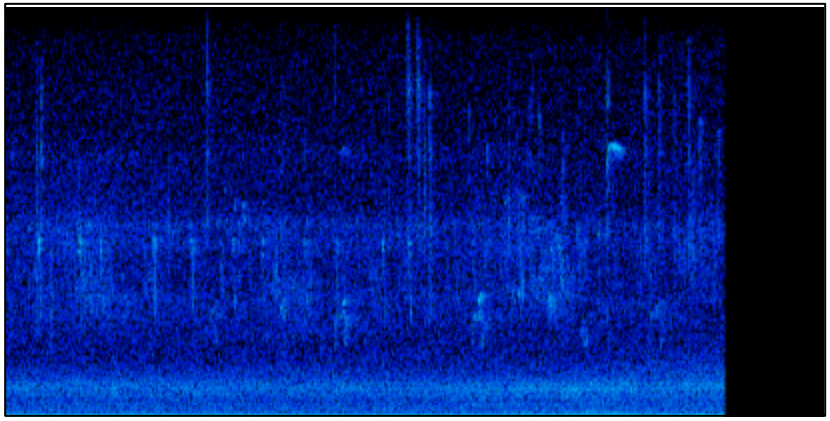
- LTER Portal/Single Sign-on ☑
- Security ☑
  - Authentication
  - Access Control
- Audit Logging ☑
  - Data Access
    - Intellectual Rights Protection
    - Data Use Policy
  - Resource Access
- High Performance Computing Resources ☑
- Data Storage Resources ☑
- Collaboration Tools ☑

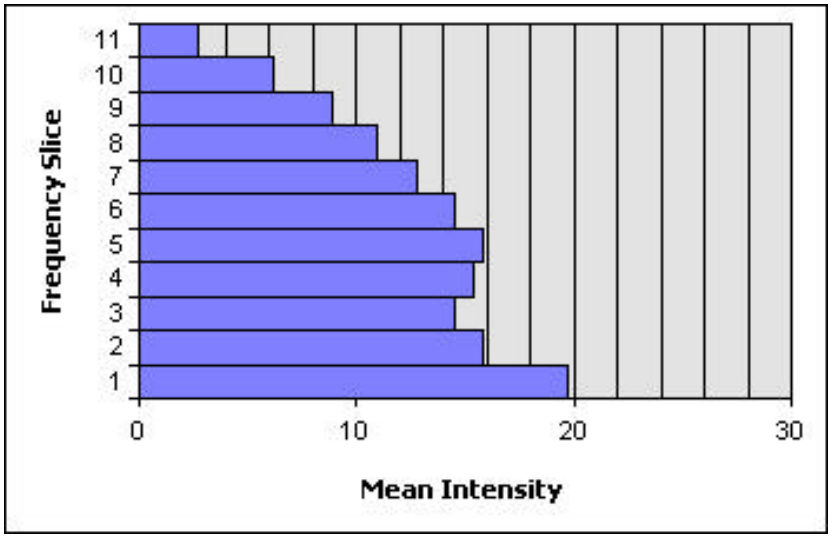
Pilot Study Experiment "Acoustic Pattern Matching"

- Environmental Acoustic Monitoring
  - Embedded microphone
  - Sampling every 30 minutes for 30 seconds
  - Conversion to *.wav file format

Pilot Study Experiment "Acoustic Pattern Matching"

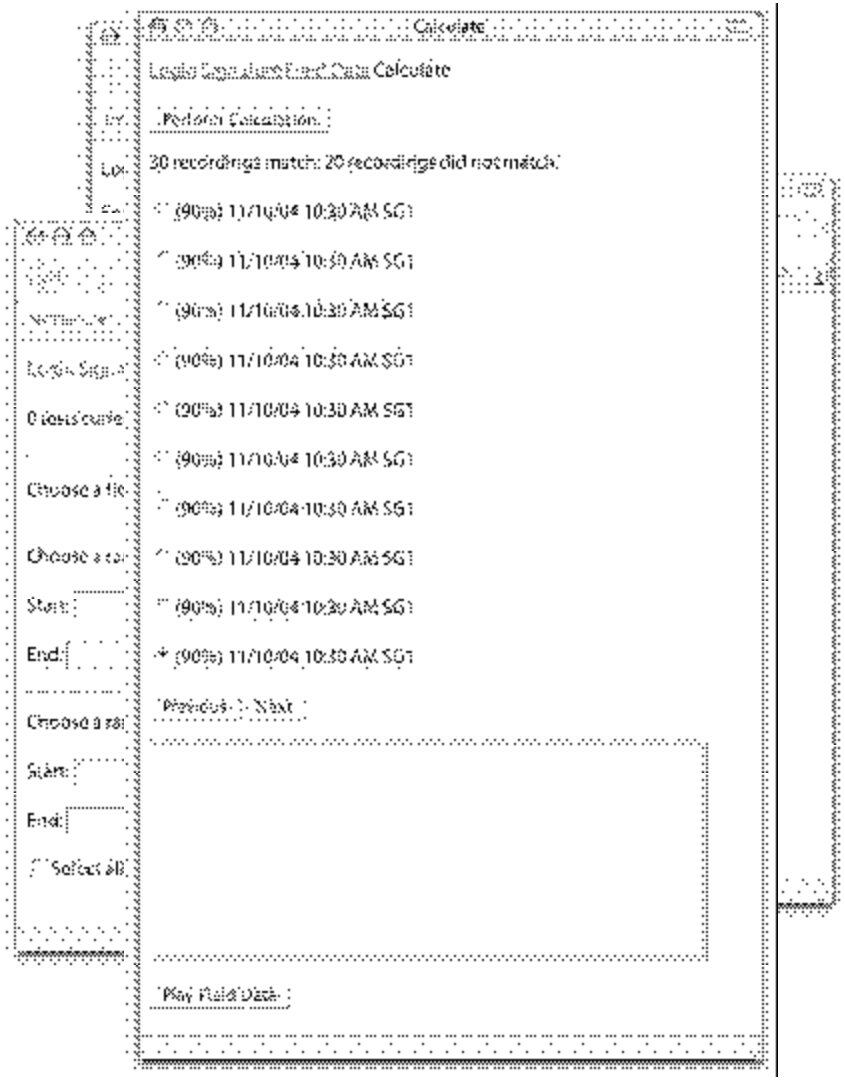
- Ornithologic (bird) pattern matching
  - DSP/Matlab workbench
  - Match unknown dataset to an known signature





User Scenario

- User logs onto portal (single sign-on)
- Selects known signature file(s)
- Performs data search via Metacat for unknown acoustic file(s)
- Performs acoustic matching on HPC; match probability identified to user

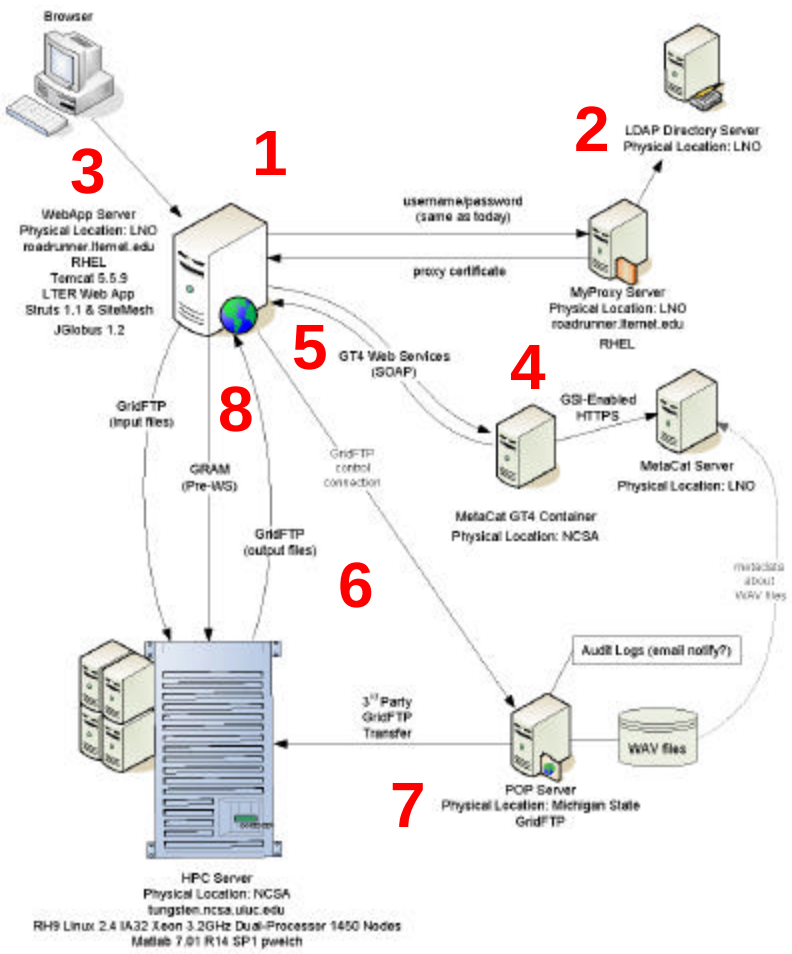


Architectural Components

- Web application - noname (J2EE using struts and jglobus)
- Certificate Management – MyProxy/LDAP (now being used by TeraGrid)
- Data Transfer - GridFTP
- Job Management - GRAM (pre-WS)
- GSI Authentication – Metacat GT4 Container
- Metadata Database - Metacat 1.5 modified for GSI (Java servlet)
- Data Discovery - Metacat Query Inteface (Java servlet using struts framework)
- DSP Analysis - Matlab

Infrastructure Scenario

1. Portal accepts user credentials and passes to MyProxy
2. MyProxy validates thru LDAP and issues certificate back to Portal
3. Portal accepts acoustic signature selection by user
4. Portal accepts acoustic query; passes query and certificate to GT4/Metacat; request logged
5. Metacat returns query results to Portal; return logged
6. Portal accepts selection of unknowns by user; passes certificate and data transfer request to data POP and HPC
7. POP and HPC accepts certificate; POP, via GridFTP, transfers data to HPC
8. HPC performs matching; returns results back to Portal



Current Status

- ***March 2005***
  - Generate detailed work-flow and application(s) specified
  - Generate detailed project plan
  - Identify components, services, and APIs
  - Identify roles and responsibilities
  - Generate detailed architecture
  - Identify site, personnel, and data involved
  - Team meeting at NCSA
- ***April 2005***
  - System and User Environment configuration
  - Software/hardware development and integration
- ***June 2005***
  - Software/hardware development and integration
- ***July 2005***
  - Release of functioning prototype
  - Prototype testing and evaluation
  - Site deployment of software/hardware
  - Site integration of software/hardware
  - Site testing and evaluation
- ***August 2005***
  - Release of working system
  - System testing and evaluation
  - Preparation of demonstration
- ***September 2005***
  - Demonstration

IM meeting, Montreal ‘05

LTER

Network Information System

EML status of LTER sites

Iñigo San Gil

Aug 5th 2005

EML status of LTER sites

• More than 90% of the LTER sites have implemented the EML standard.

• About 75% have made EML metadata available at centralized servers. These "Metacat" servers have harvested over 3,000 EML documents from LTER sites.

"Harvested" here is loosely defined as: at least a few site metadata sets of low content placed in the server. Also, the site has a good plan to place all legacy data in, and with specific plans to enrich EML, if appropriate.

• 2/3rds of the LTER sites are *in*; factoring out the two new LTER sites ~ 75% of the sites are harvested

How about the 25% ?

FCE	:	About to be harvested to an high EML level
KNZ	:	Has very rich EML metadata
HFR	:	Has implemented EML level 3ish. (To complete: Entity table)
PAL	:	Focus on site reorganization process, long term plans
CDR	:	Working on EML implementation and harvest as we speak
BNZ	:	Positively intrigued about it
JRN	:	On the verge of posting level 5 EML for 70% of datasets

EML status of LTER sites

Some more brief site notes:

- BES has a proven concept for the harvesting, only a few files are in
- SGS has a model to convert the metadata, and tried it on about ten files.
- ARC, PIE, NWT, CWT, LUQ, KBS, HBR have all legacy metadata harvested to a respectable richness level (Tier 2-3.5).
- MCM and NTL have some metadata harvested, and I would like to work with them to see where do they plan to go from here.
- The remaining 6 LTER sites; AND, CAP, GCE , SBC, SEV and VCR have harvested their metadata with all sorts of details (rich EML). All but SEV have all their legacy metadata in the Metacat servers, and SEV will follow soon.

EML status of LTER sites: Some numbers

Site	Harvesting Since	TIER level	Harvested sets	% EML/ % Harvested
AND	Jun ’05	5	124	100 / 100
ARC	Apr ’05	2 ½	1585	100 / 100
BES	Apr ’05	1	2	5 / 5
CAP	Aug ’04	5	30	100 / 25
CWT	May ’05	2 ½	190	100 / 97
GCE	Apr ’04	5	245	100 / 100
HBR	Jul ’04	4	112	100 / 100
KBS	Aug ’04	3	40	100 / 100
LNO	Jan ’05	2 (a few 4)	360	100 / 100
LUQ	May ’05	3 ½	96	100 / 100
NTL	Apr ’05	3 ½	43	? / ?
NWT	Jun ’05	2 ½	139	100 / 100
PIE	Jul ’05	2 ½	111	100 / 100
SEV	Jul ’05	3 ½	46	~20 / 20
SGS	Jun ’05	2	9	10 / 10
VCR	Jul ’05	5	63	100 / 100
SBC*	’04	5		100 / 100

*SBC harvests to NCEAS metacat directly. NTL has all but GIS legacy metadata compliant

BES converted 100% of their metadata, still curating.

EML status of LTER sites: Some numbers

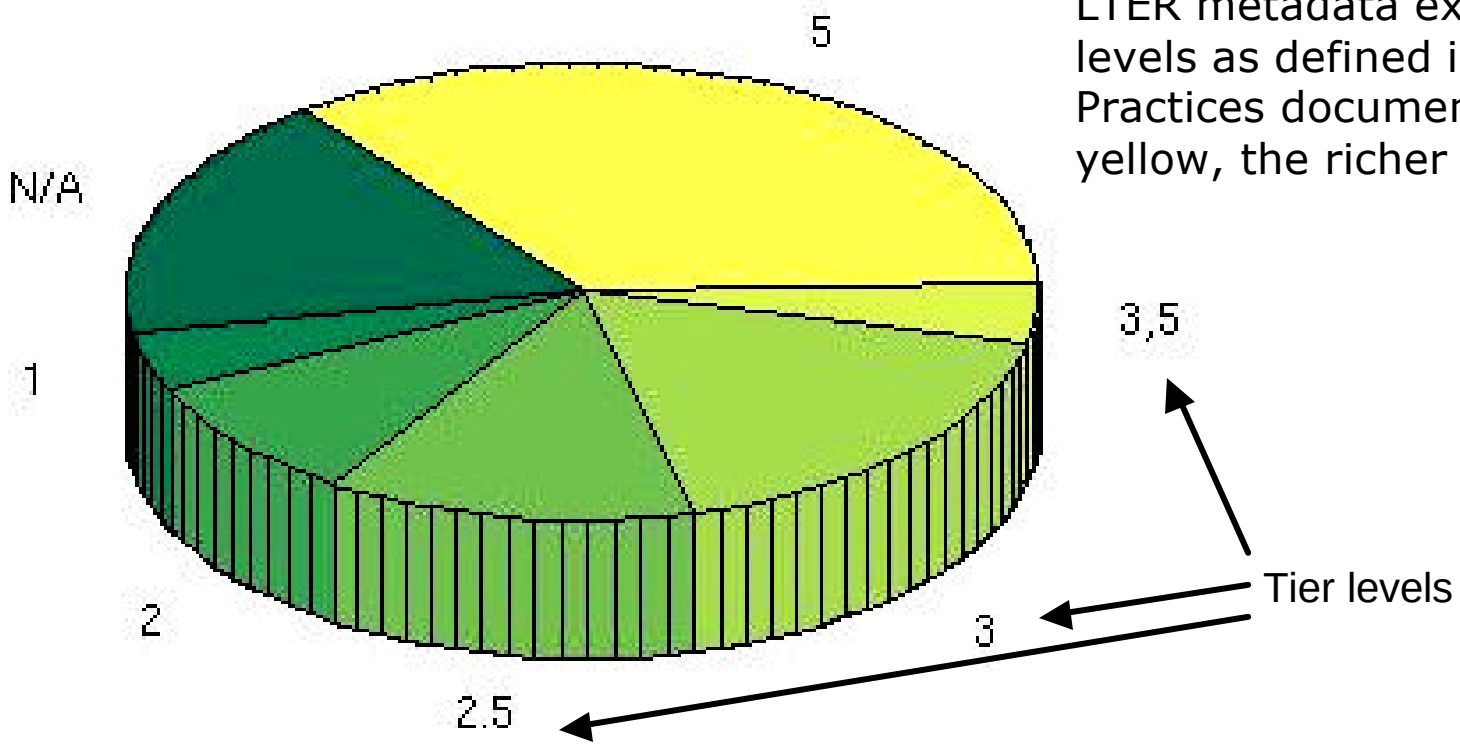
Site	Harvesting Since	TIER level	% EML / % Harvested
BNZ	N/A	--	-- / --
CDR	N/A	--	-- / --
FCE	N/A	5	100? / 0
HFR	N/A	3+	100 / 0
JRN	N/A	--	-- / --
KNZ	N/A	5	100 / 0
PAL	N/A	--	-- / --
MCR	New Site		
CCE	New Site		

BNZ has plans to start EML implementation on September 05. CDR is currently working on EML standard adoption
FCE and KNZ are about to harvest.

EML status of LTER sites: Tier levels

Pie chart of LTER EML metadata documents completeness*

What see here is an estimate of the richness of the standardized LTER metadata expressed as Tier levels as defined in the EML Best Practices document: the more yellow, the richer the metadata



*estimate values. Certain inconsistencies noted

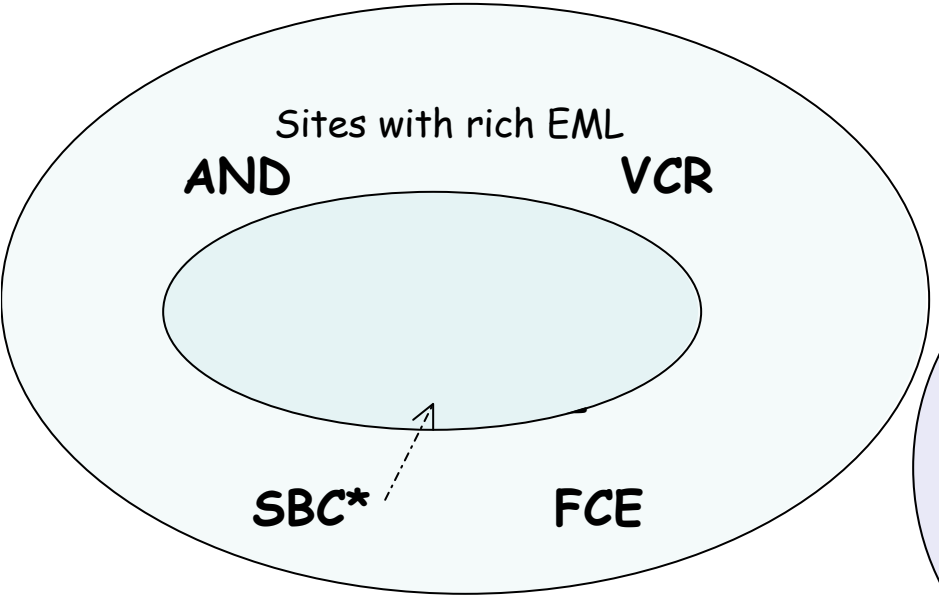
Evolution of EML status of LTER sites

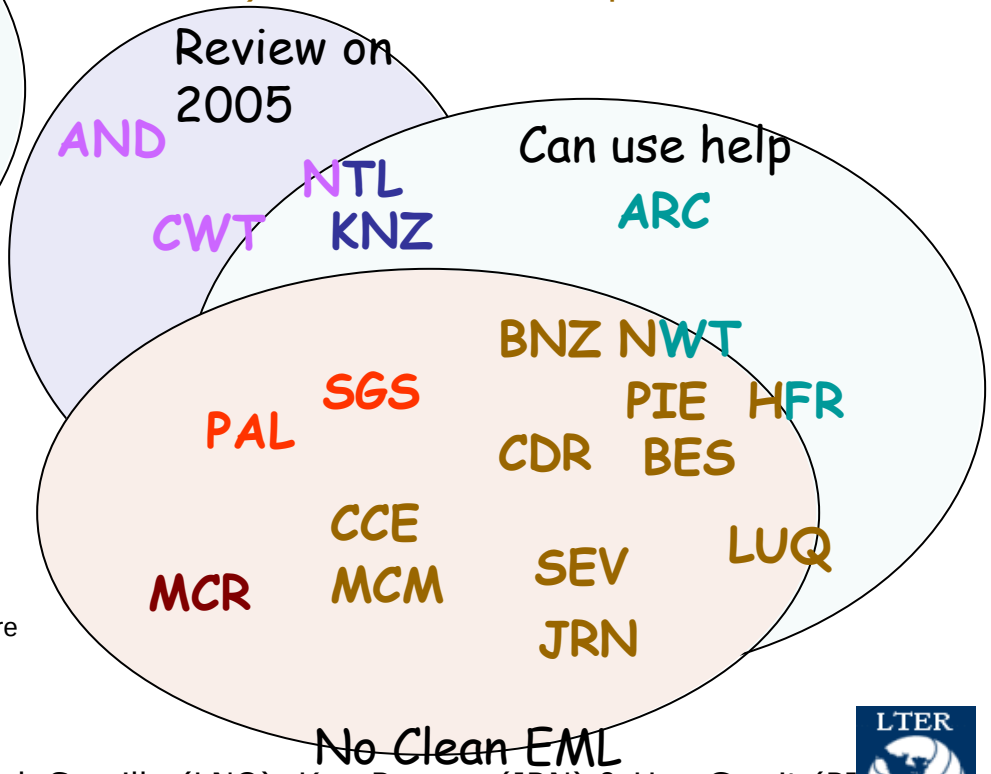
Overview of EML implementation as of March 05

Mar 2005

In these three diagrams you can see three criteria to categorize the sites needs for EML help:

1) Under review this year
2) Asked for help
3) None or not complete EML





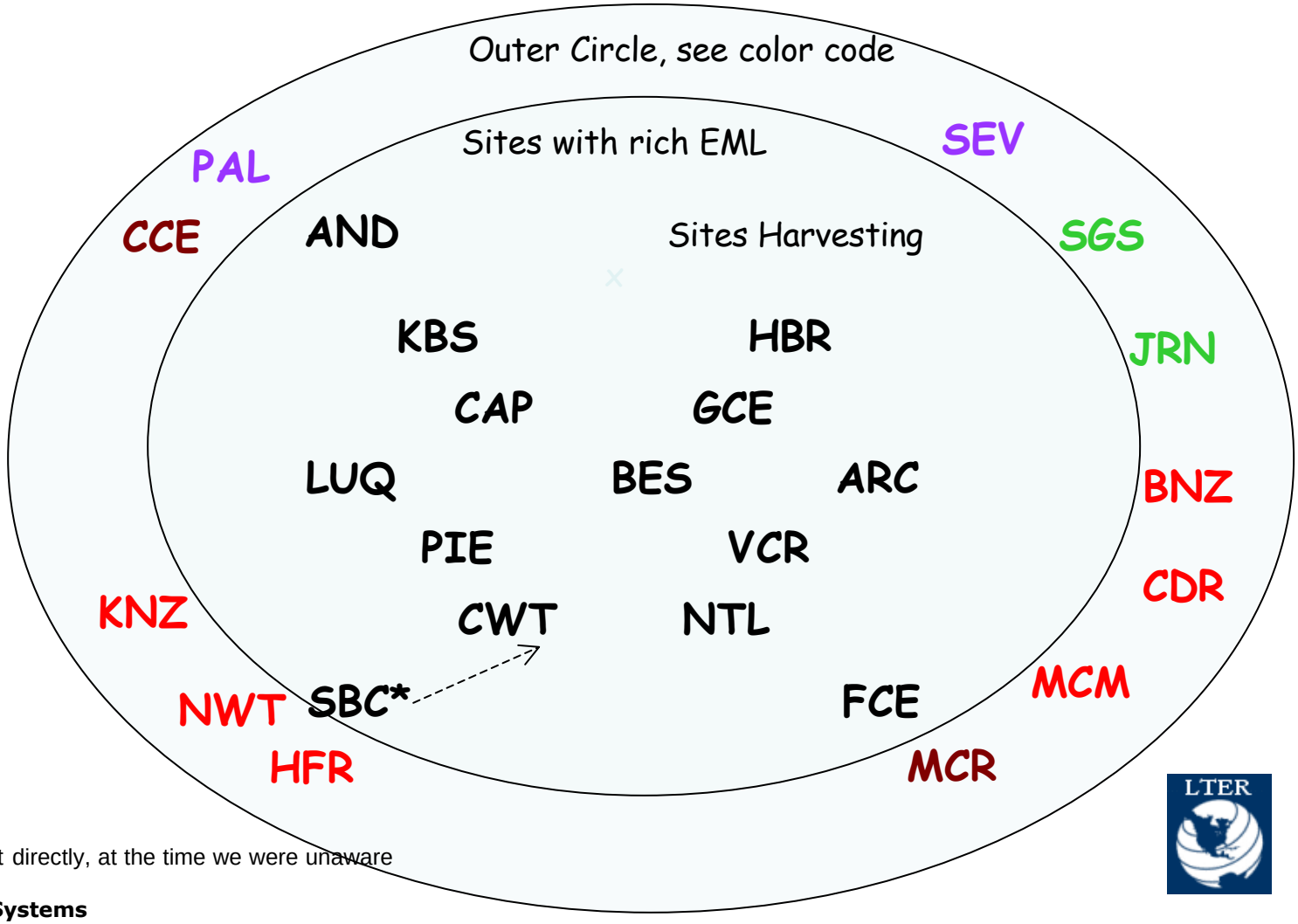
*SBC harvests to NCEAS metacat directly, at the time we were unaware

Diagrams based on EML survey conducted by Mark Servilla (LNO), Ken Ramsey(JRN) & Hap Garrit (PI and also the 2005 LNO survey.

Evolution of EML status of LTER sites

- Visited, work in progress
- Possible site visit
- Contacted
- New Site

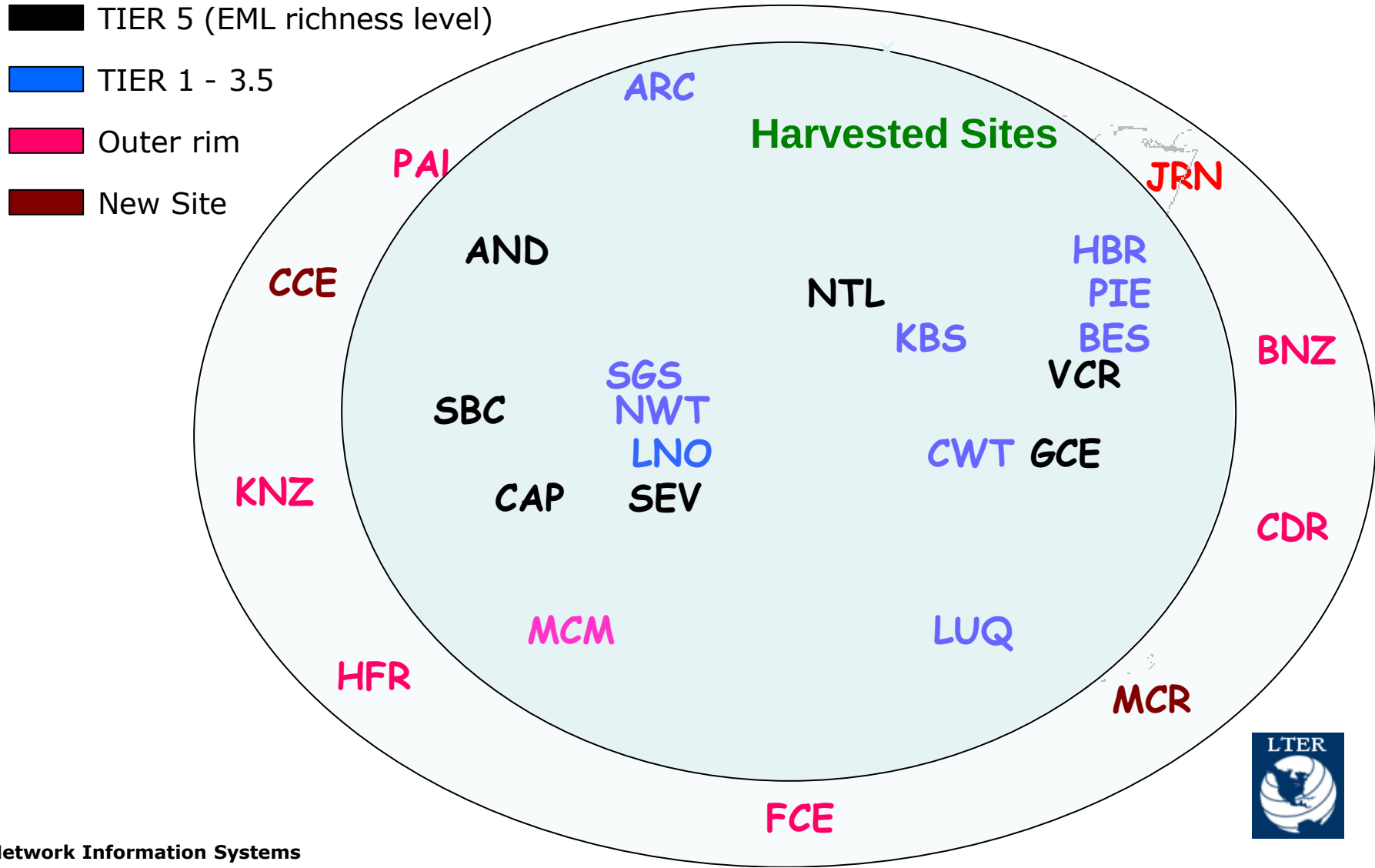
Jun 2005



*SBC harvests to NCEAS metacat directly, at the time we were unaware

Evolution of EML status of LTER sites

Aug 2005



EML status of LTER sites

Jimmy Awards

remarkable achievements

Early EML adopter, implementer	GCE, CAP
Volume of documents	ARC
Richest EML	AND, GCE
Most patient IMs	PIE, ARC
Taxonomic	NTL, NWT
Creative	CAP
Thorough and enthusiastic	LUQ
Promising	BES
Best network team	LTER: all YOU!!

Long Term Ecological Research
Network Information System

The Long Term Ecological Research Network Information System Strategic Plan

Presented for the NISAC committee by:
James W Brunt
Associate Director for Information Management
LTER Network Office

Network Information System: Vision

- an infrastructure for confederation and research collaboration
- depending on site data repositories
- comprised of information technologies and the information products
- primarily serves the LTER scientific community and collaborators,
- provides a “portal” to LTER data products for the broader scientific community, natural resource managers, policymakers, and the general public.

Network Information System: Vision

- will achieve it's goals by implementing shared standards, software tools, training, and support
- Integrate and provide compatibility across the Network of sites, institutions, and researchers.
- balance development of short-term, priority products with development of long-term, generic solutions and improved network-wide infrastructure.

Network Information System: Vision

- Additionally, viewed as conceptual framework that enhances flows of information between data, synthesis and knowledge about ecological systems
- depending on the science and Information Management (IM) participants at the site and network levels for implementation,
- relying on and contributing to informatics expertise and IM systems outside of LTER.

NIS Conceptual Model

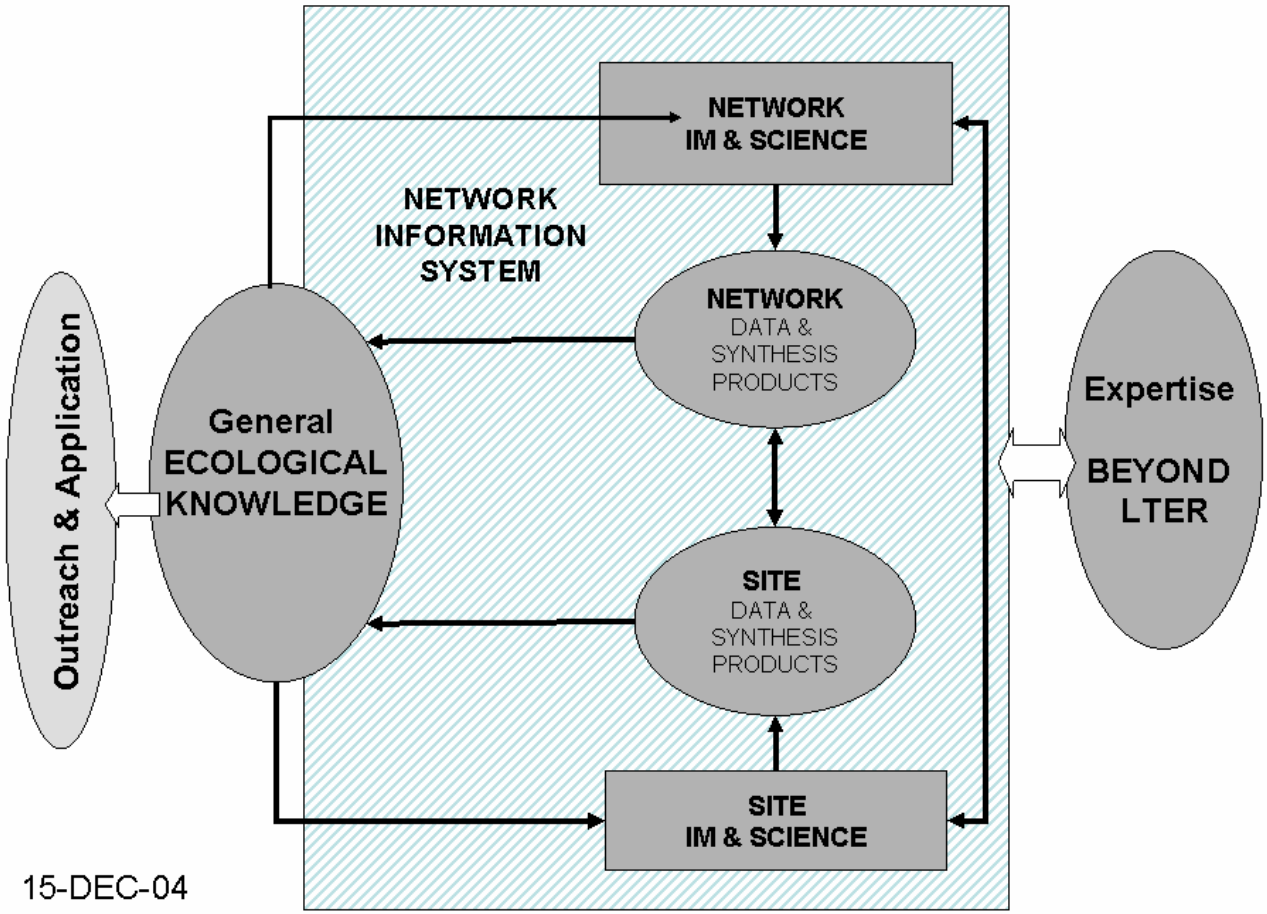


Figure 1 – Schematic diagram of NIS conceptual framework.

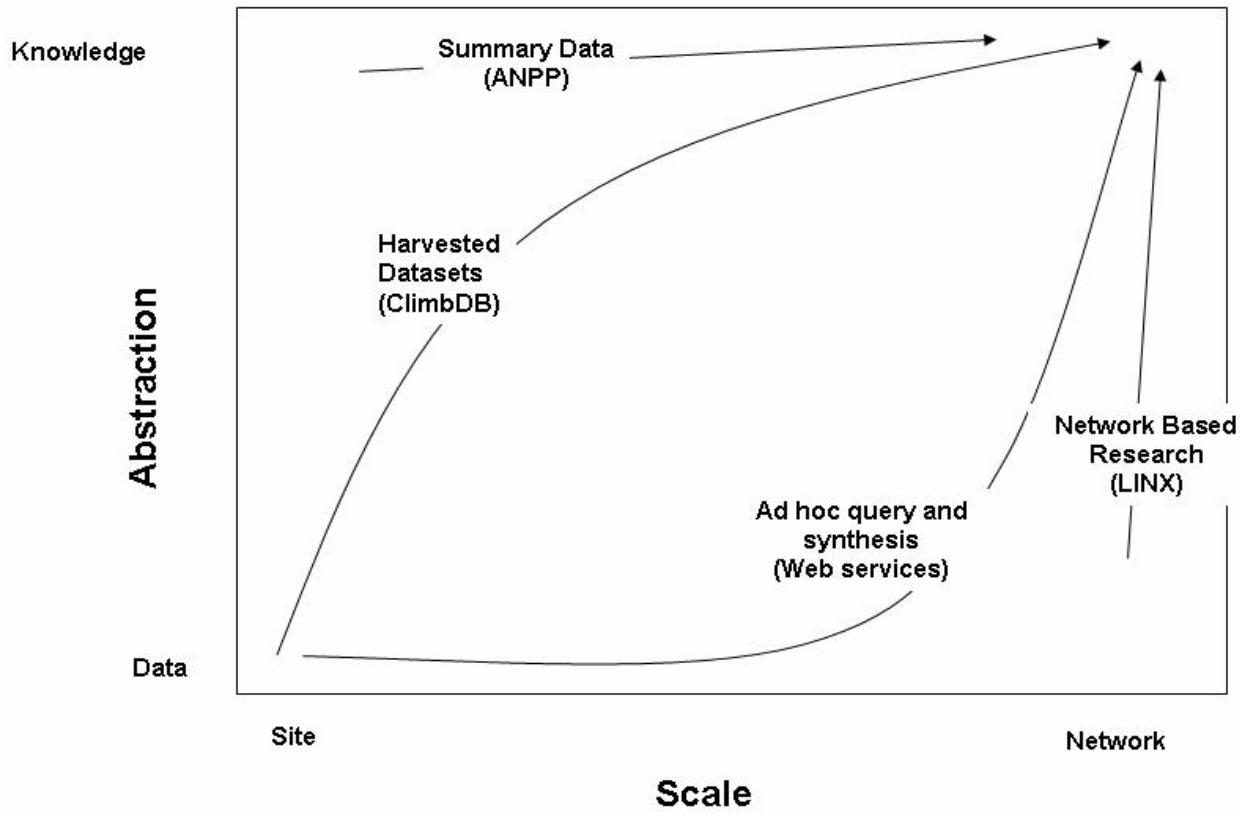
Network Information System: Mission Statement

- The mission for the LTER Network Information System (NIS) is to provide the Information Management and Information Technology infrastructure to facilitate and promote advances in collaborative and synthetic ecological science at multiple temporal and spatial scales.

Network Information System: Goals

1. Data: Increase data quality through standard approaches
2. Synthesis: Increase data available for synthetic activities
3. Knowledge: Increase knowledge discovery through synthesis

NIS synthesis approach diversity



Data: Increase data quality through standard approaches

Strategies:

- Increasing the quality of data and metadata by adopting, developing, and implementing standards.
- Increasing support for local site information management and information technology to accommodate new standards, infrastructure, etc.
- Providing training, guidance, and workshops on new information technology and information management to sites.

Synthesis: Increase data available for synthetic activities

Strategies:

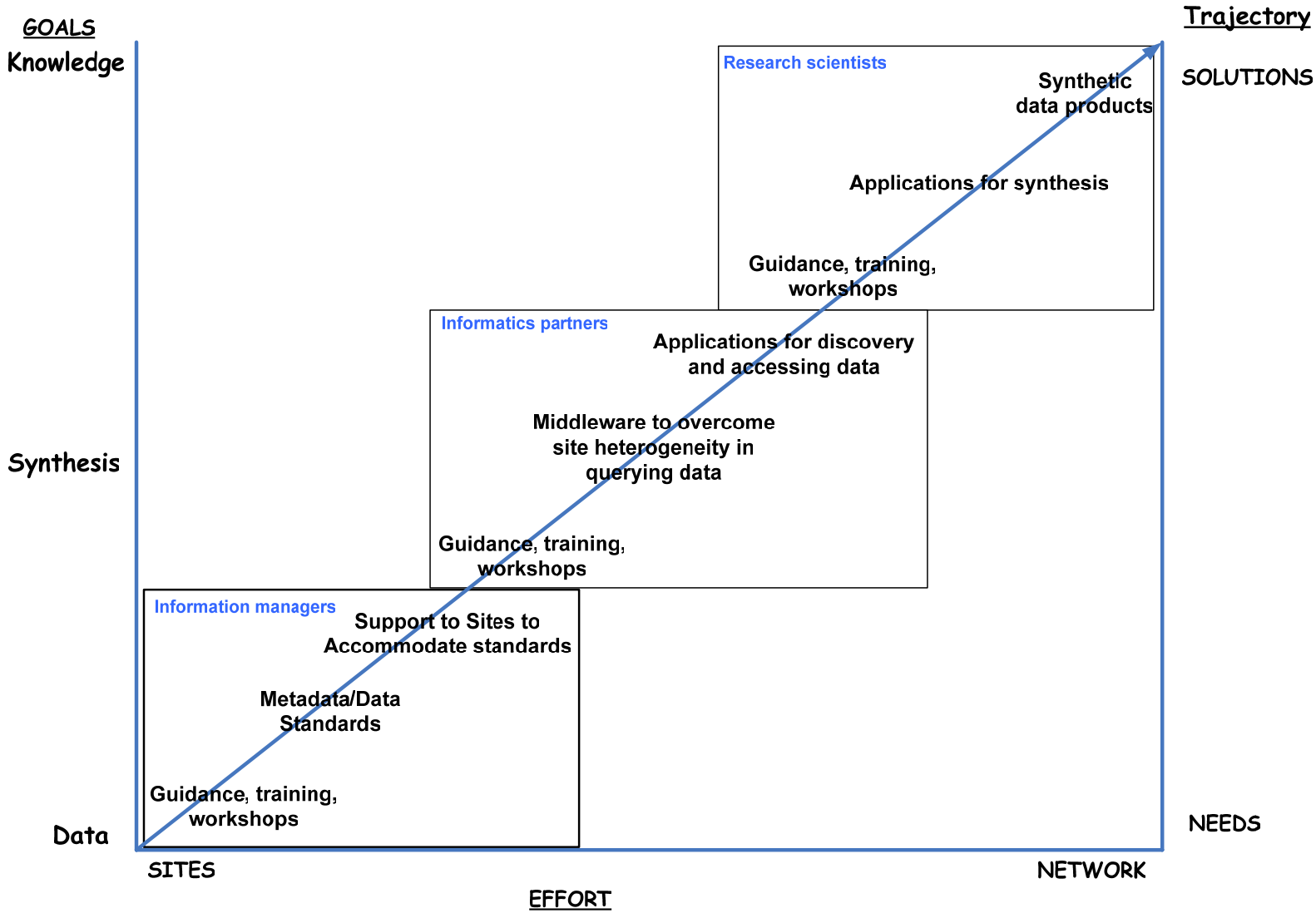
- Increasing the use of available data by developing and deploying basic applications and interfaces for discovering and accessing LTER data.
- Overcoming site heterogeneity issues by developing, adopting, or adapting middleware to enable seamless Get and Query access to site data.
- Supporting workshops for network developers to integrate specific information technologies.

Knowledge: Increase knowledge discovery through synthesis

Strategies:

- Providing education and training to researchers on NIS applications.
- Supporting exploitation of LTER information products through advanced application development.
- Supporting the creation of Network-based synthetic information products through application development, scientific collaboration, and product-oriented workshops.

NIS Trajectory w/ current major strategies



Network Information System: Implementation and Process

Roles

- **NISAC** – NISAC is responsible for planning, promotion, and evaluation of NIS activities.
- **LNO** – LNO will lead the design, development, deployment, and support of NIS infrastructure.
- **CC** – The coordinating committee will approve and endorse the goals and strategies elaborated in the NIS strategic plan.
- **SITE** – Sites will implement Network standards, provide access and timely updates to sites research data and metadata in formats agreed upon.
- **IMC** – The Information Management Committee will review NIS strategic plan action items directly affecting the sites and communicate comments to NISAC. The **IMEXEC** will survey, accumulate, and communicate status of progression of sites through the tiered trajectory to NISAC.

LTER Network Information System Advisory Committee

- Henshaw, Donald - AND (chair)
- Benson, Barbara – NTL
- Boose, Emery - HFR
- Brunt, James - LNO
- **Gage, Stuart - KBS**
- **Harmon, Mark - AND**
- **Kratz, Timothy - NTL**
- McCartney, Peter - CAP
- Michener, William - LNO
- **Peters, Debra - JRN**
- **Ross, Robin – PAL**
- Servilla, Mark - LNO
- Vande Castle, John - LNO
- Waide, Robert - LNO

NIS Strategic Plan: Next Steps

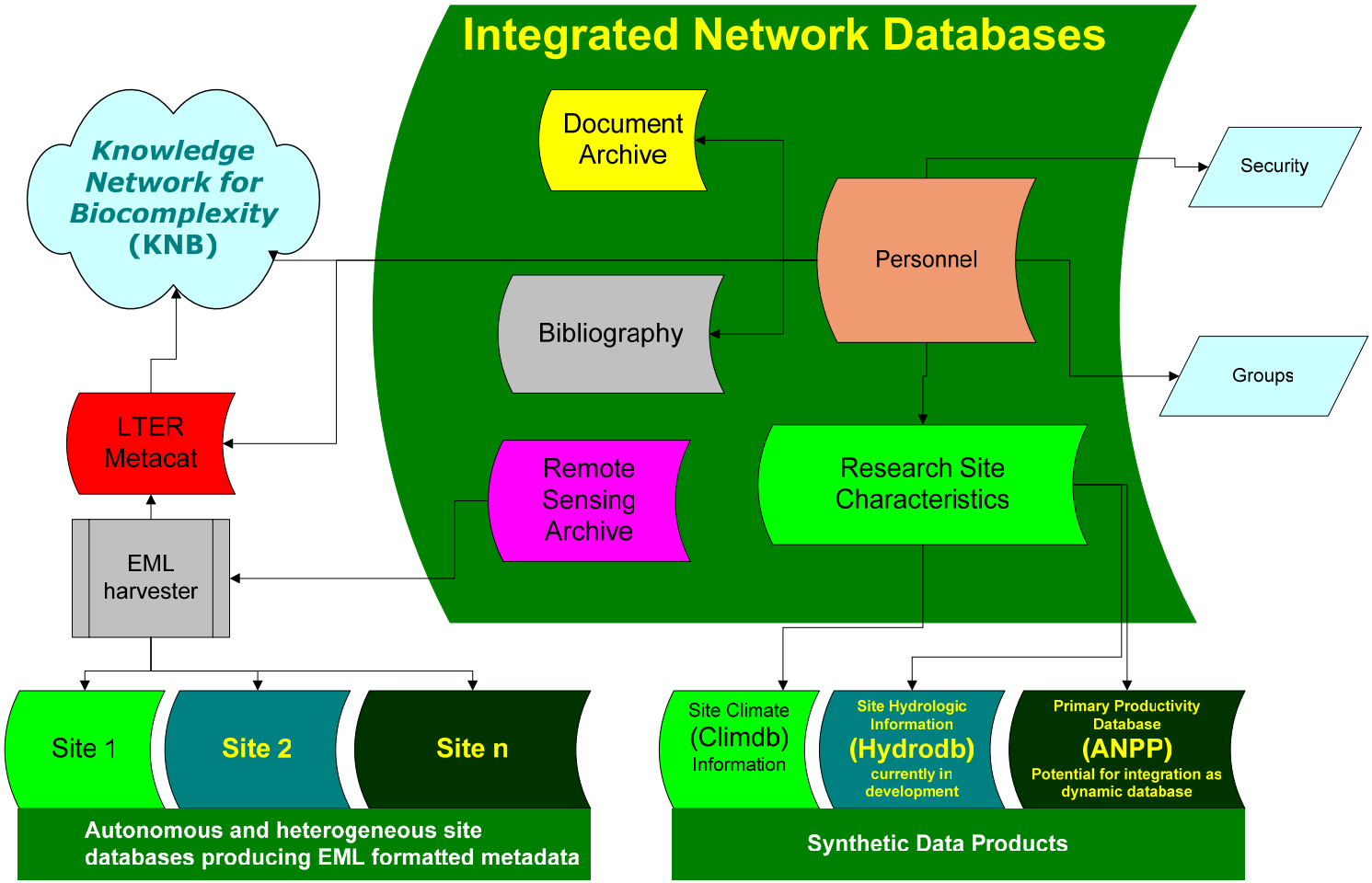
- Current version presented to IMC August 4, 2005
- IM commentary period until August 19, 2005
- NISAC final revisions by September 9, 2005
- To the coordinating committee for approval September 20-21, 2005
- Implementation plan process begins October 2005

Long Term Ecological Research
Network Information System

Status of LTER Network Databases

James W Brunt
Associate Director for Information Management
LTER Network Office
&
others

Network Databases



Site Characteristics

- Presented at Santa Barbara CC
- Pre-populated at LNO
  - Site profiles
  - Call for updates in 2004
- Need for documentation of locations, themes
- Need more data in:
  - Research Locations
  - Research Themes
  - Additional interfaces

Document Archive

- Redesigned in 2005 based on site surveys
  - Final documents
- Reorganized into content management system
  - Includes moderated upload
- Next Steps
  - Richer Metadata
  - Search Interface

LTER

Remote Sensing Archive

- Discovery metadata registered in LTER metacat, Spring 2005
- Complete metadata for one TM time-series as an example and to gain RS/EML experience.
- Next Steps
  - Publication
  - Richer Metadata for all
  - SRB registration/replication

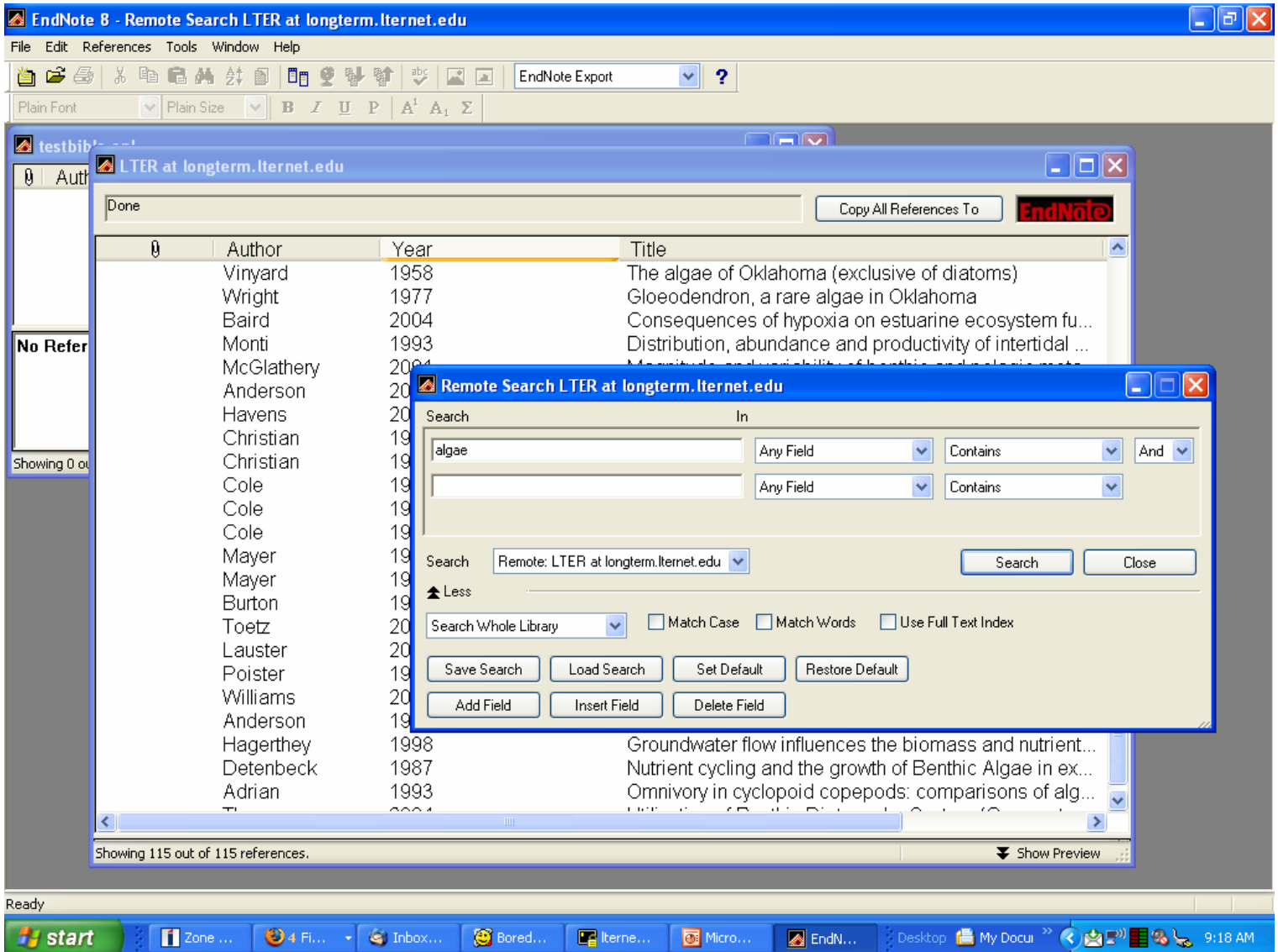
Personnel

- Currently implementing revised data model to accommodate web service updates
- Call for updates late fall
- Hard copy directory in Spring ’06
- Next Steps:
  - LDAP reconfiguration
  - Identity management separation

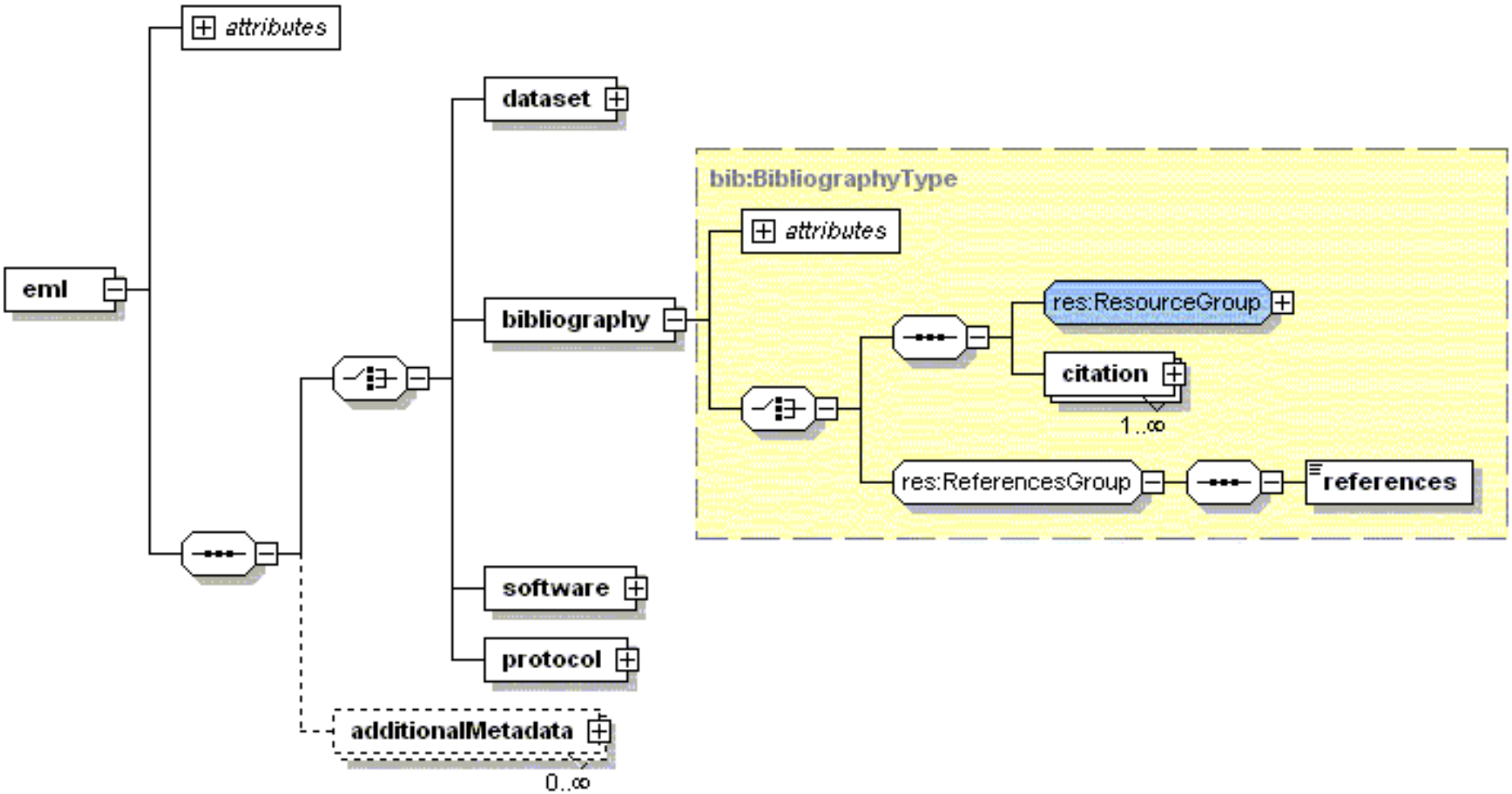
Bibliography

- Relational database and php interface in 2004
- Updates requested spring of 2005
- Participation – 80%
- Reports -
- Z 39.50 – connection information on cvs.lternet.edu
- Query interface available on cvs.lternet.edu
- Next Steps:
  - Decision on unique ID and LTER funding
  - Linkages to Personnel
  - EML connector
    - Recommend downgrade eml-literature to eml-citation
    - Recommend create eml-bibliography module (resource group plus 1-many eml-citation elements)

Bibliography



Bibliography



Climbdb
Hydrodb

ClimDB/HydroDB Database Projects - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.fsl.orst.edu/climhy/

The Mozilla Organiza... | Safari Tech Books O... | Site Characteristics ... | Ecological Metadata ... | Creative Commons | CTWatch Quarterly ...

Climate and Hydrology Database Projects (CLIMDB/HYDRODB)

Climatological and Hydrological Data Access

Please note that these pages work better using the most recent versions of Internet Explorer and Netscape.

Welcome to CLIMDB/HYDRODB, a centralized server to provide open access to long-term meteorological and streamflow records from a collection of research sites.

Please review the Data Access Policy before using the data.

- Contributors

	View All	LTER	USFS	USGS
Sites	◉	○	○	○
Stations	◉	○	○	○
Variables	◉	○	○	○

- Data, Plots, and Downloads
- Metadata Reports
  - Complete Site Report (PDF)
  - By Category Report (HTML)

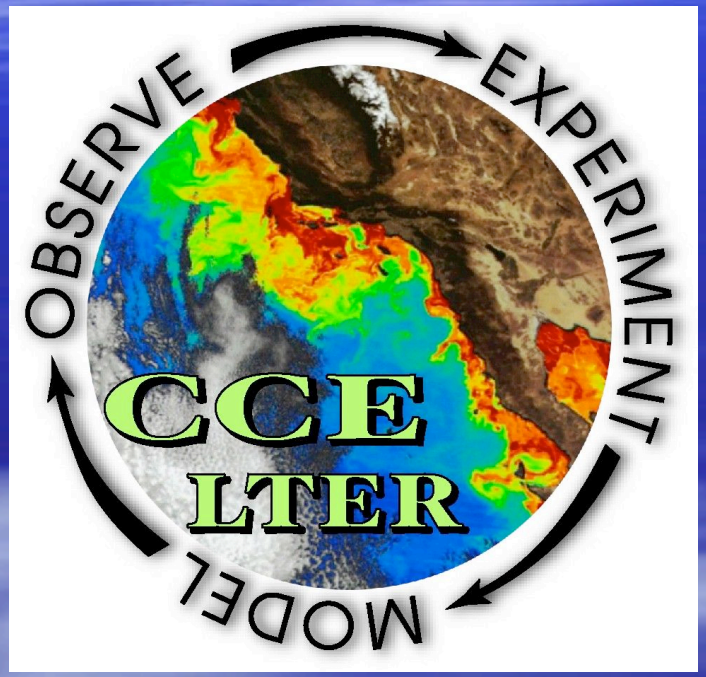
- General Description

Done

start | network-databases.ppt | EndNote 8 - Remote ... | 3 Firefox | Desktop | My Documents | 12:52 PM

California Current Ecosystem (CCE) LTER Site

A coastal upwelling biome



Lead PI: Mark Ohman
Information Manager: Karen Baker

Ocean Informatics Team:

- **Jerry Wanetick - Computational Systems Center Director**
- **Lynn Yarmey - Programmer-analyst, Metadata**
- **Shaun Haber - Database/Web Designer, HCI**
- **Mason Kortz - Web/DB Designer, Systems Admin**
- **Jim Wilkinson - CalCOFI Data Manager**

Based at the **Scripps Institution of Oceanography**

LNO's Thomas McOwiti joins CCE first cruise

Many institutions and research goals under CCE umbrella

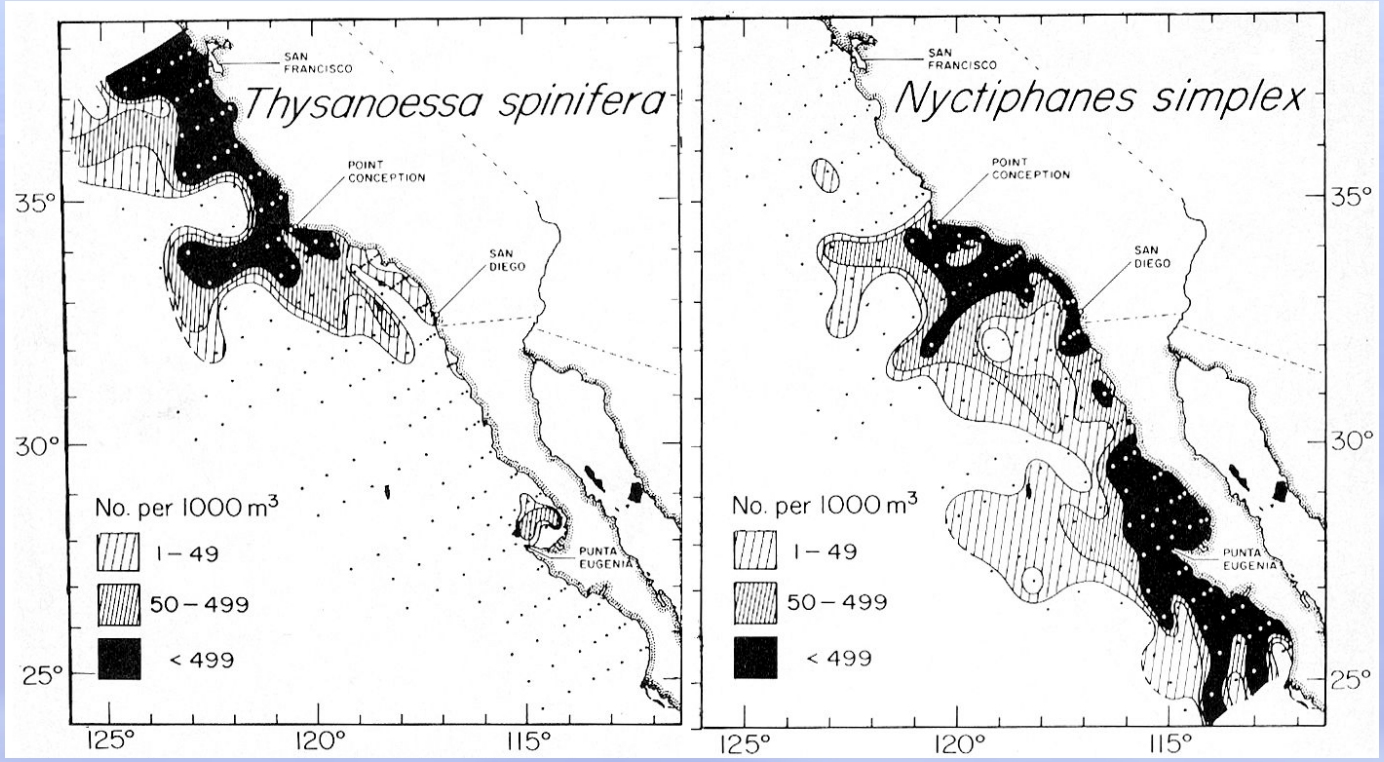
CCE LTER Program Participants

Name	Role	Institution	Interests
Mark Ohman	Lead PI	SIO	Mesozooplankton Ecology
Lihini Aluwihare	Co-PI	SIO	Dissolved Organic Matter
Karen Baker	Co-PI	SIO	Information Management
Katherine Barbeau	Co-PI	SIO	Iron Geochemistry
David Checkley	Co-PI	SIO	Mesozooplankton & Ichthyoplankton
Peter Franks	Co-PI	SIO	Biophysical Modeling
Ralf Goericke	Co-PI	SIO	Phytoplankton Ecology
Michael Landry	Co-PI	SIO	Food-Web Structure and Function
Art Miller	Co-PI	SIO	Physical Oceanography; Modeling
Greg Mitchell	Co-PI	SIO	Remote Sensing and Bio-optics
George Sugihara	Co-PI	SIO	Nonlinear Modeling
Farooq Azam	Associate	SIO	Bacteria/Microbial Food Webs
Steven Bograd	Associate	PFEL	Physical Oceanography
Ron Burton	Associate	SIO	Molecular Probes for Protists
Dan Cayan	Associate	SIO	Atmospheric Physics
Teresa Chereskin	Associate	SIO	ADCP Currents
Emanuel DiLorenzo	Associate	Georgia Tech.	Biophysical Modeling
David Hyrenbach	Associate	Duke/PRBO	Seabird Ecology
Cheryl Peach	Associate	BAS	Education and Outreach planning
Sharon Franks	Associate	CA COSEE	Education and Outreach planning
Brian Palenik	Associate	SIO	Microbial Diversity
Dan Rudnick	Associate	SIO	Mesoscale Ocean Physics
Christian Reiss	Associate	SWFSC	Planktivorous Fishes
Ken Smith	Associate	SIO	Deep-sea Benthic Ecology
Bill Sydeman	Associate	PRBO	Seabird Ecology; Marine Mammals
Elizabeth Venrick	Associate	SIO	Phytoplankton Floristics

SIO = Scripps Institution of Oceanography
PFEL = Pacific Fisheries Environmental Laboratory
Duke = Duke University
SWFSC = Southwest Fisheries Science Center/U.S. National Marine Fisheries Service
BAS = Birch Aquarium at Scripps
Georgia Tech = Georgia Institute of Technology
PRBO = Pt. Reyes Bird Observatory
CA COSEE = California Center for Ocean Sciences Education Excellence

Why focus on this part of the California Current System?

- A Biogeographic Boundary Region



Brinton 1981, Haury et al. 1986

- Large and dynamic range of biological and physical ocean conditions
- 55 years of CalCOFI observations
- Non-linear event patterns and transitions

CCE Program Elements

- **Science Goals**

  What are the mechanisms leading to different ecosystem states in a coastal pelagic ecosystem?

  What is the interplay between changing ocean climate, community structure and ecosystem dynamics?

- **Science Methods**

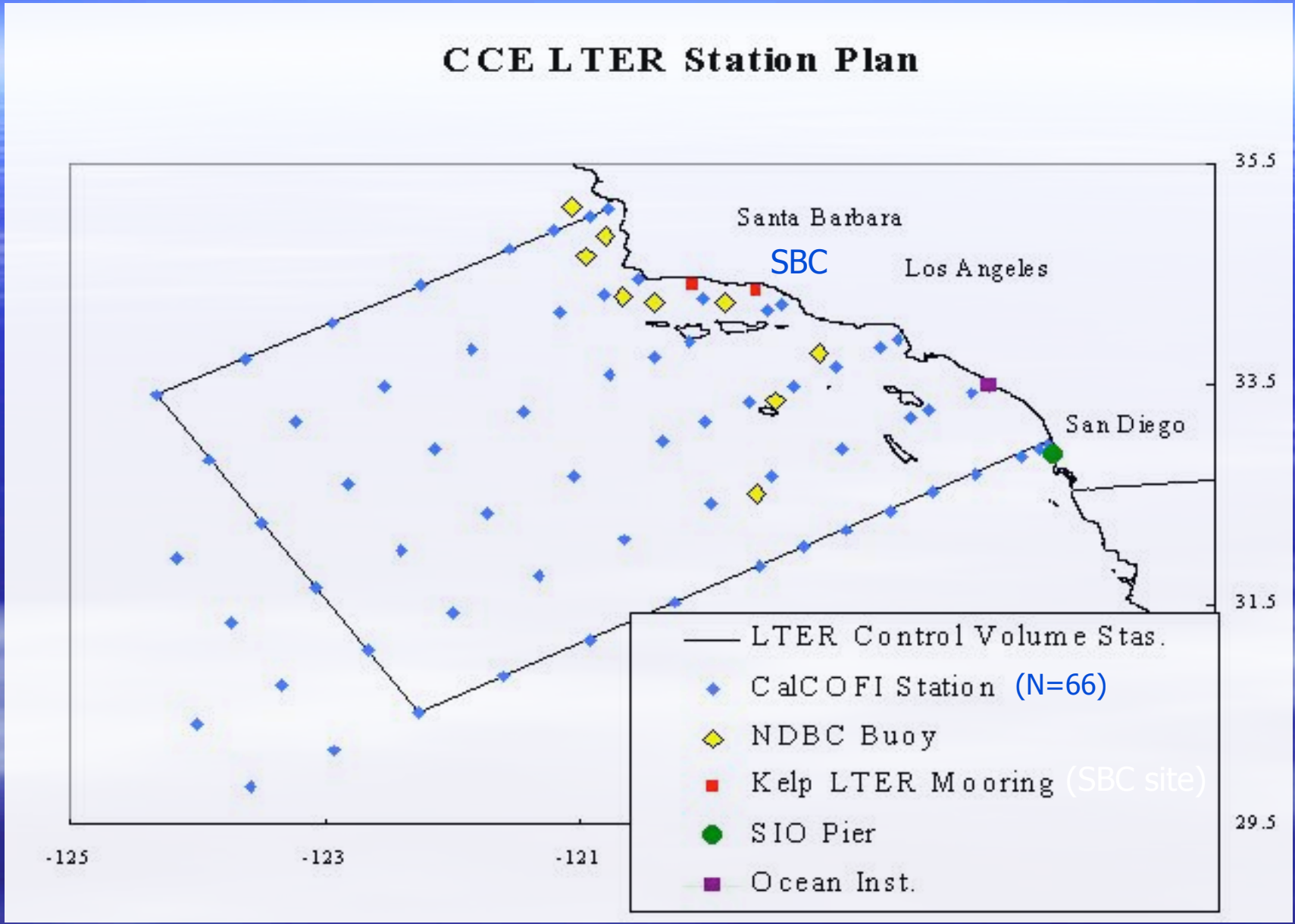
  Time Series (building on 55 year CalCOFI study)
  Physical collections
  Satellite remote sensing
  Modeling

- **Information and Data Management**

  Metadata Initiatives
  Infrastructure
  Data Capture, Curation and Access
  Data Interoperability
  Sociotechical Informatics

- **Education and Outreach**

Spatial Grid



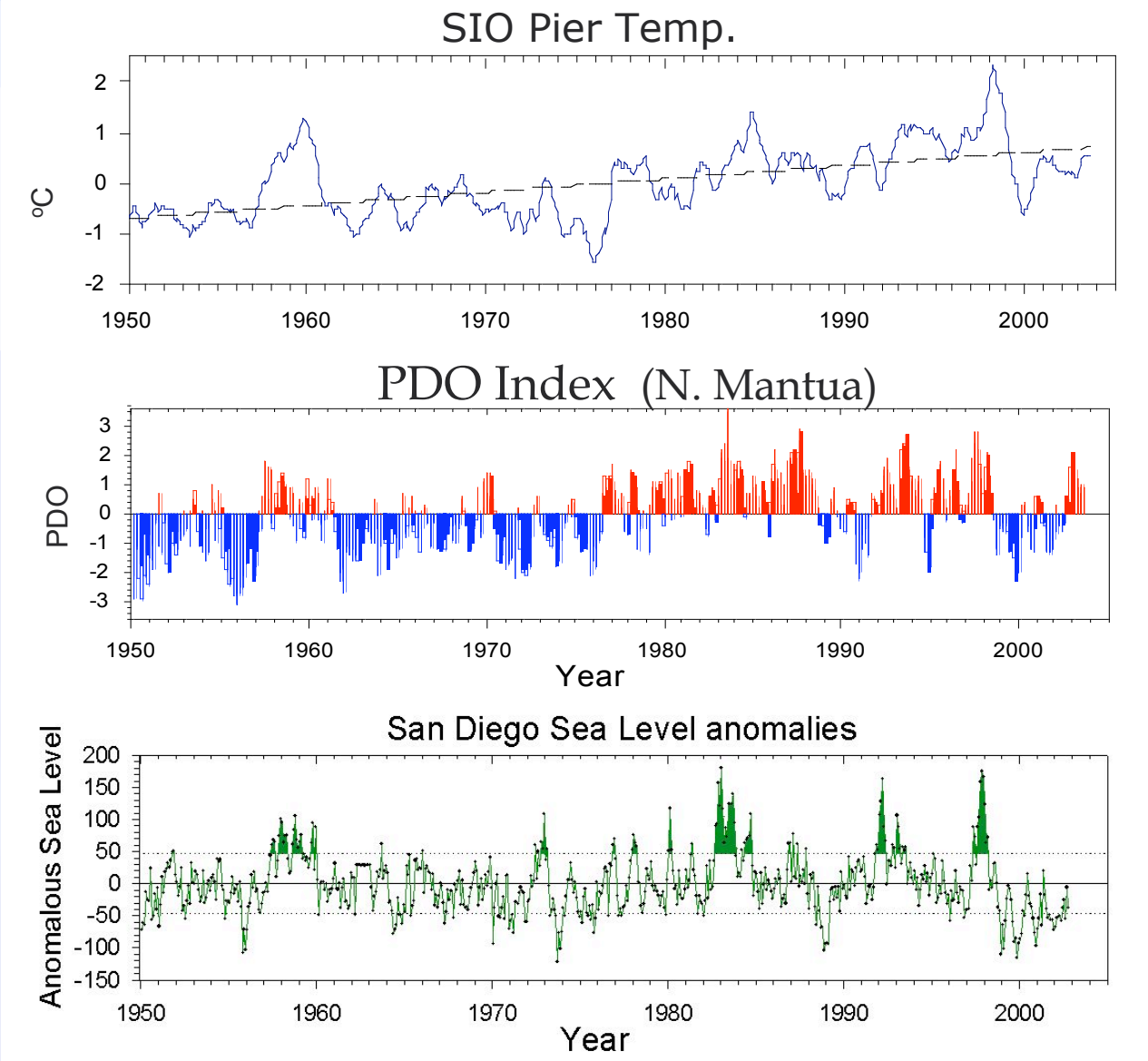
Time Series

Secular trend

PDO

Pacific Decadal Oscillation

El Niño



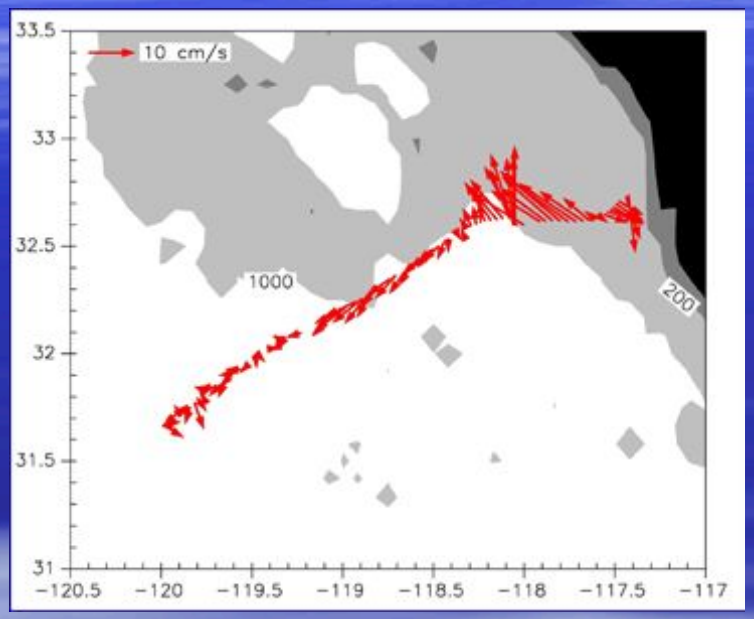
Collections

Bongo Net Deployment

SIO: Invertebrates Collection
NOAA: Fish Eggs and Larvae

Recent enhancements to our LTER site

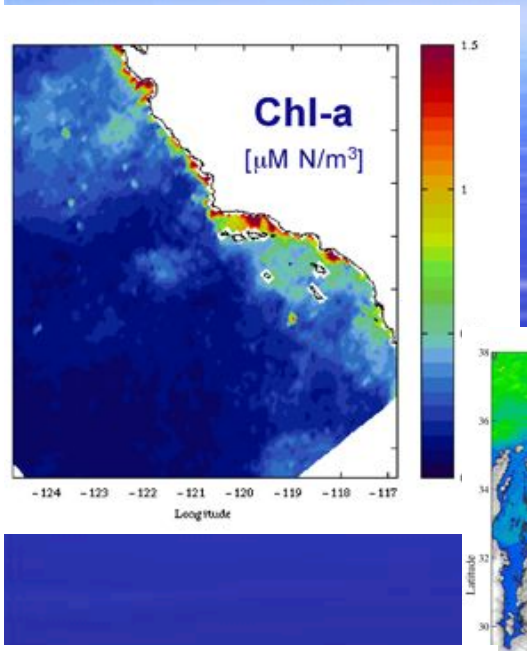
Spray gliders

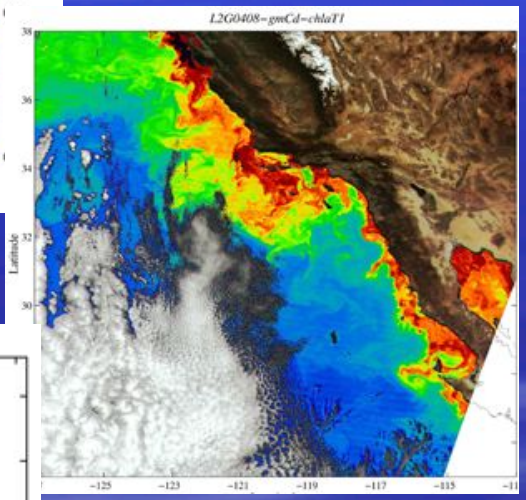


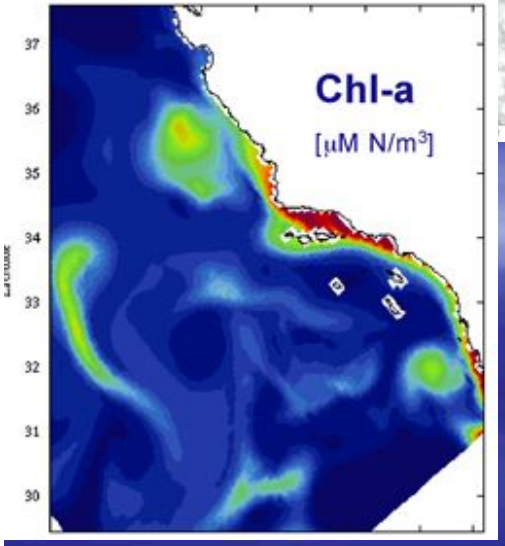
velocity, CalCOFI line 90

intended: sensors
- Wetlabs fluorometer
- Sonntek 500 kHz ADCP
- CTD

Benefits of a large study area







Remote Sensing

Exploiting **spatial** differences as an analogue of **temporal** change

Ecosystem Model

Information and Data Management Work Flow

Field → **Design** → **Field**

Design Studio
Design Teams
Working Groups
Community Meetings

Manually recorded data, one field log for each measurement type

Augmenting and automating field data management practices

CCE Information/Data Management:

Design process underway

- Ocean Informatics Environment
  - -open source centered
  - -new server installed: iOcean
  - -local partnering initiated:
    - CalCOFI, PaCOOS, SCCOOS, Palmer LTER
- Infrastructure:
  - -distributed computing
  - -design team approach
- Ongoing activities:
  - -data policy, metadata (EML)
  - -survey, collaborative tools
- Human Social Dynamics NSF project:
  - -natural/human/information interfaces

Education and Outreach

Beth Simmons

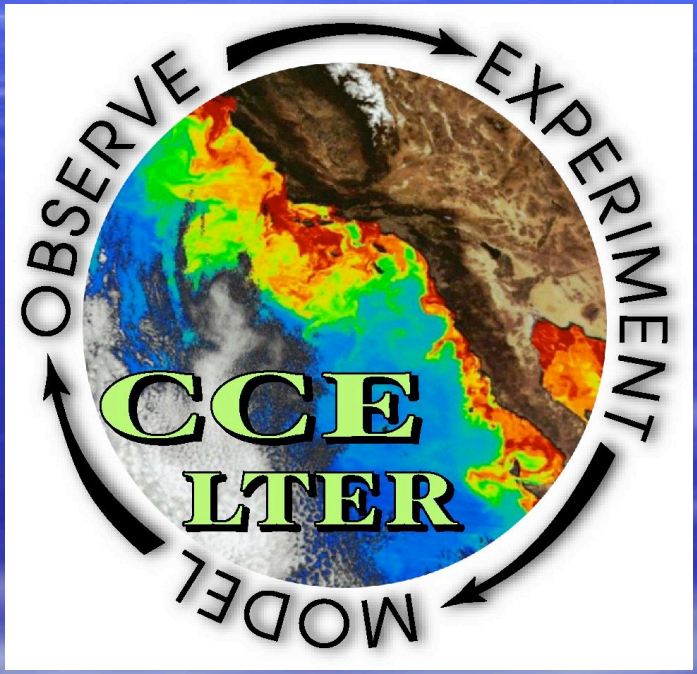
Education Outreach Coordinator

Mary Cerullo

Children's Author

CCE: California Coastal Ecosystem

Information Management supporting CCE science within an Ocean Informatics Environment



http://ccelter.sio.ucsd.edu

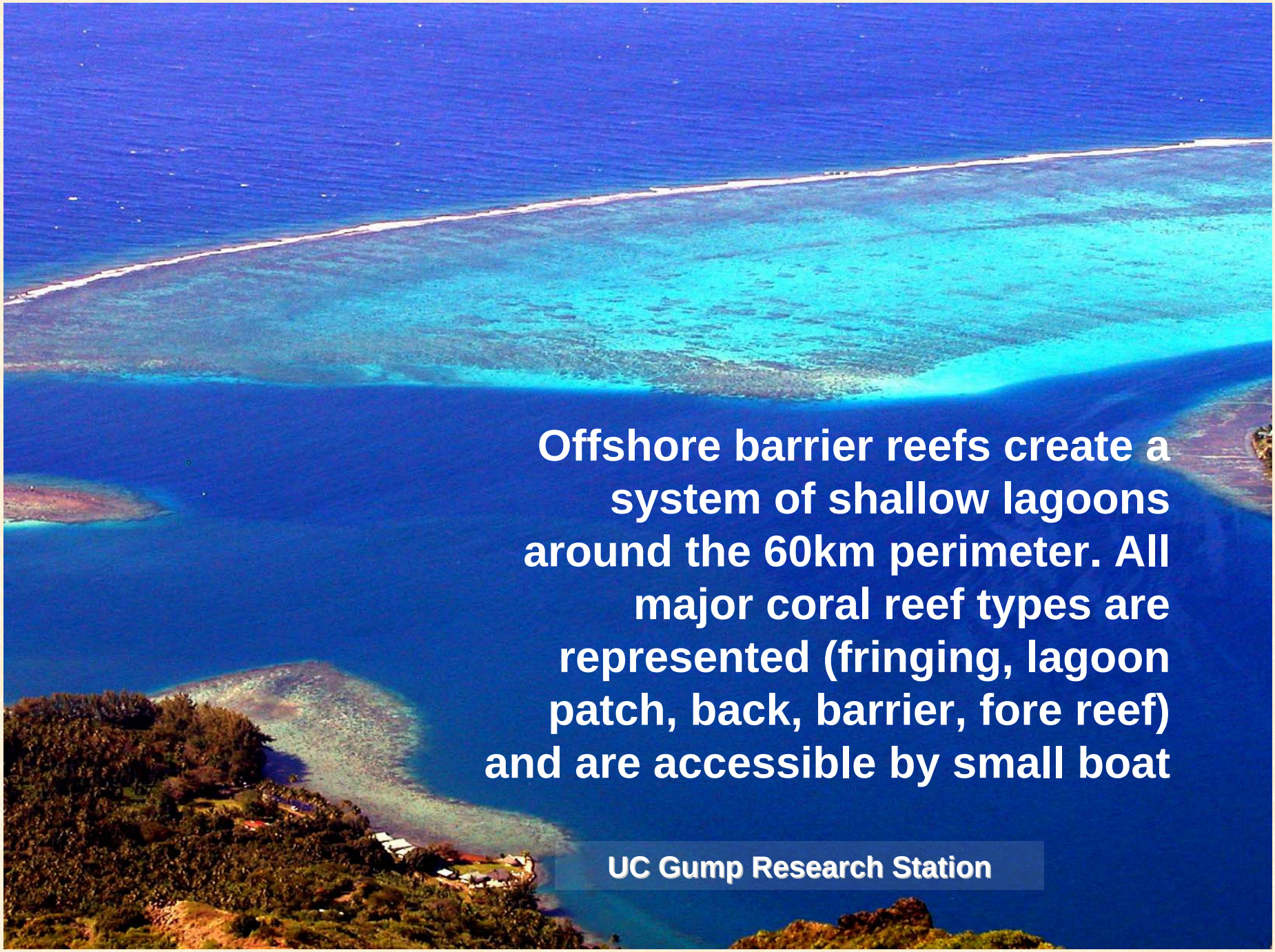
http://cce.lternet.edu

Moorea Coral Reef Long - Term Ecological Research Site (MCR LTER)

PIs: Russ Schmitt, Sally Holbrook,

Pete Edmunds, Bob Carpenter


50N
170W
160W
150W
140W
130W
100W
40N
NORTH AMERICA
North Pacific Ocean
30N
Tropic Of Cancer
20N
10N
0
10S
20S
30S
40S
Tropic Of Capricorn
South Pacific Ocean


Offshore barrier reefs create a system of shallow lagoons around the 60km perimeter. All major coral reef types are represented (fringing, lagoon patch, back, barrier, fore reef) and are accessible by small boat
UC Gump Research Station


Gump facilities include:
• Dormitories and bungalows
• Laboratory with flow-through sea water system
• Dock, launch ramp, small boats
• SCUBA compressor and dive locker

Gump Station Overview

- **1981** Established on property donated by Richard Gump, to the University of California, Berkeley
- **1982-1987** Infrastructure construction using donations from the Gordon and Betty Moore Foundation
- **1986** Six researchers begin studies at Gump Station, with nine additional in 1987
- **2004** Moorea Coral Reef LTER established, first field season, 2005

Close collaborators with the CRIOBE field station at neighboring Opunohu Bay (est ~1970)

Monitoring Component

A. Measure decadal trends in ecosystem attributes & environmental drivers

B. Provide a contextual basis for process - oriented studies

C. Meet the needs for comparative analyses within LTER network

Moorea Coral Reef LTER

Research Program: Dynamics of a Physically – Forced System

Science Themes

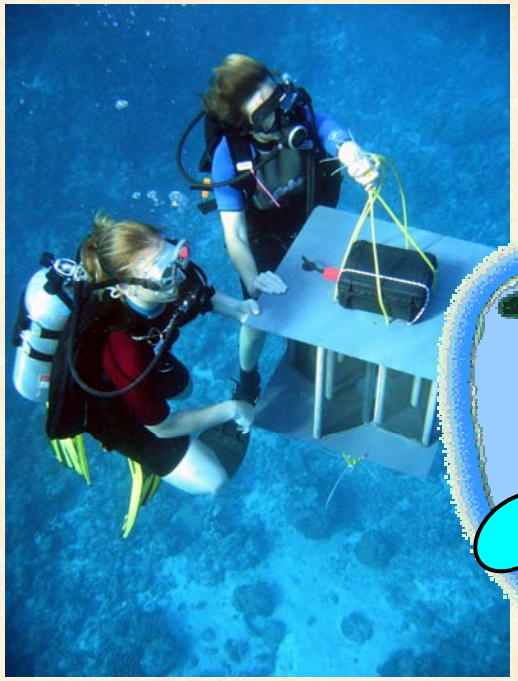
1. **Biological bases for variation in ecological performance of stony corals (the foundational group)**

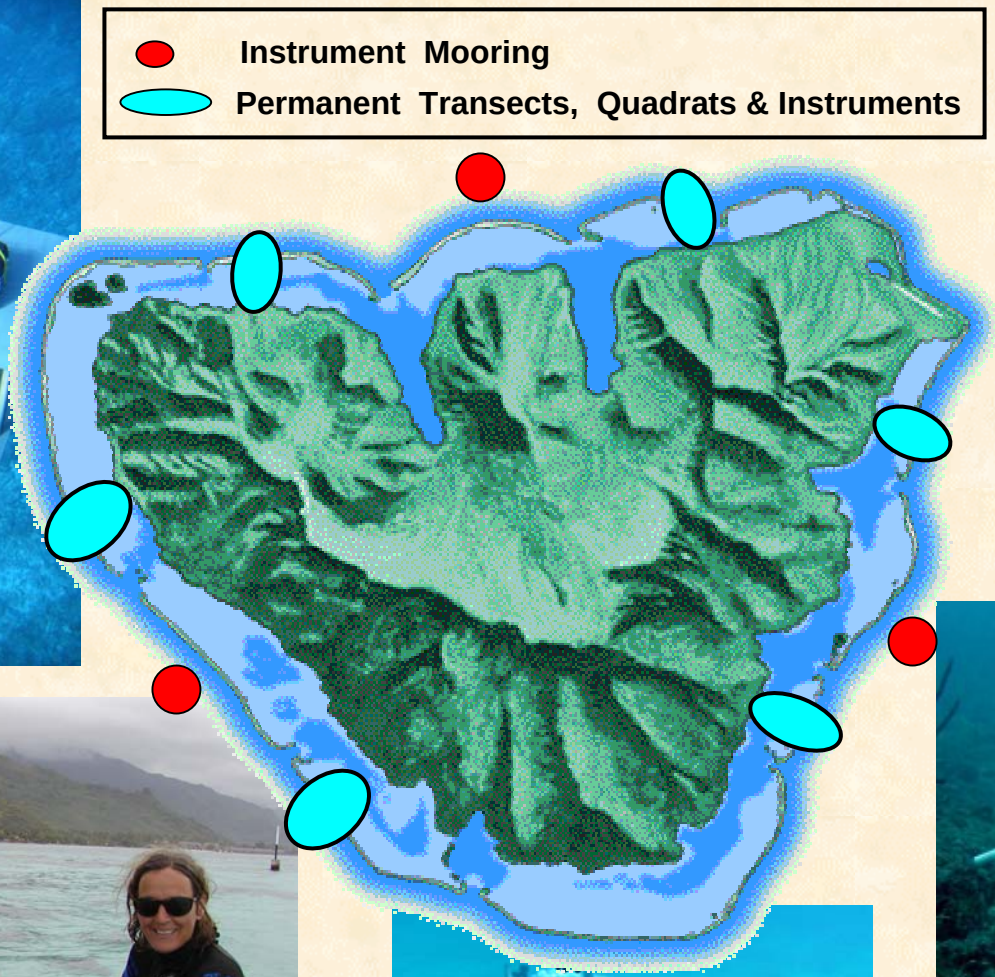
2. **Population dynamics & regulation (stony corals, fishes, algae & herbivores)**

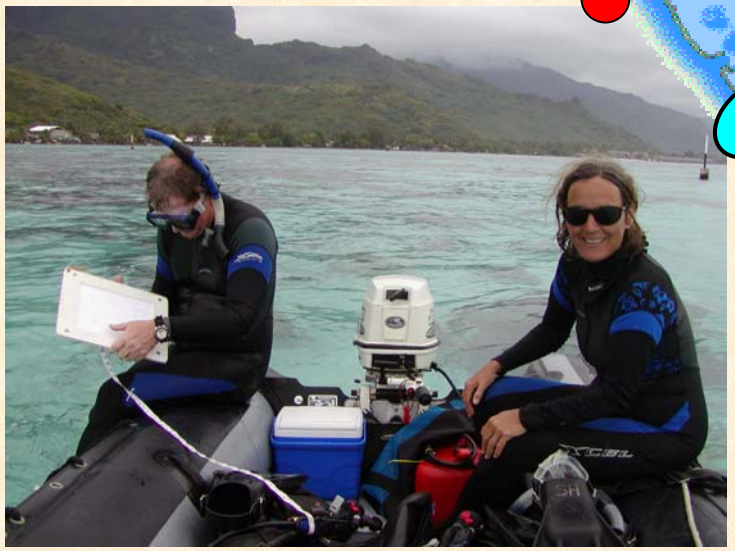
3. **Food web & nutrient dynamics**

4. **Maintenance & functional consequences of diversity**

Moorea Coral Reef LTER
Monitoring Program







SBC – MCR Collaborations

Similar Sampling and Analysis Methods
Shared investigators and staff

Cross-Site Comparisons of Reef Ecology
Tropical coral reef (MCR) vs. rocky temperate reef (SBC)

Information Management
Common data formats, metacat data catalog, ldap, servers, software, office space, …

Systems and Services

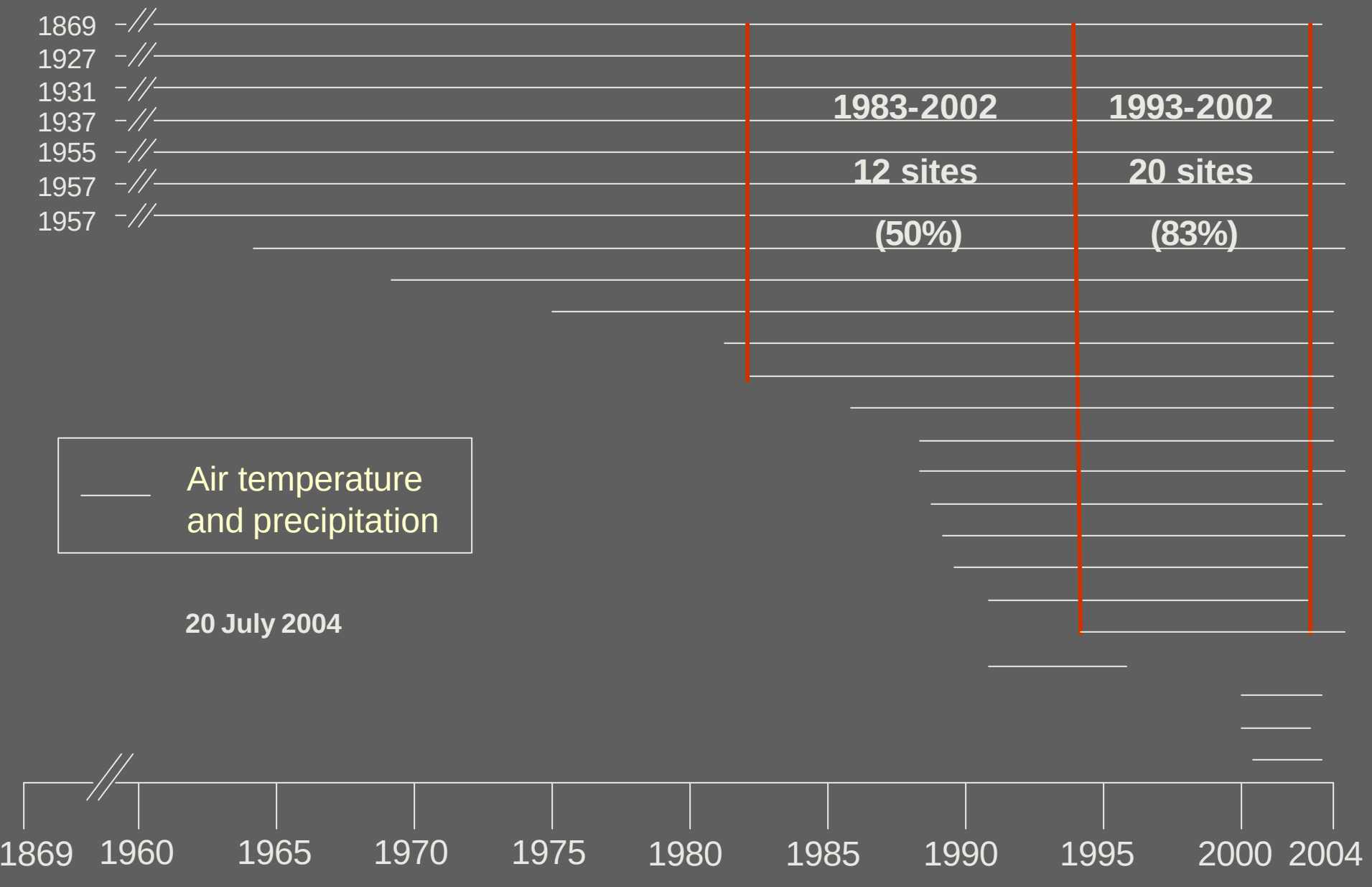
- Hardware:
  - Dual 3.6 GHz CPU IBM x336 series server
  - RedHat Enterprise Linux 4, 2.4 TB disk, 7 GB RAM
  - Distributed server backup
  - APC 3000 UPS

- Software:
  - Apache web server, Tomcat servlet engine
  - Windows SMB fileserver, Mac Appletalk fileserver
  - SSH remote login server
  - OpenLDAP personnel database
  - CVS source control
  - Postgres, SAS, Matlab, GSLIB, PERL, etc …

Shameless Plug

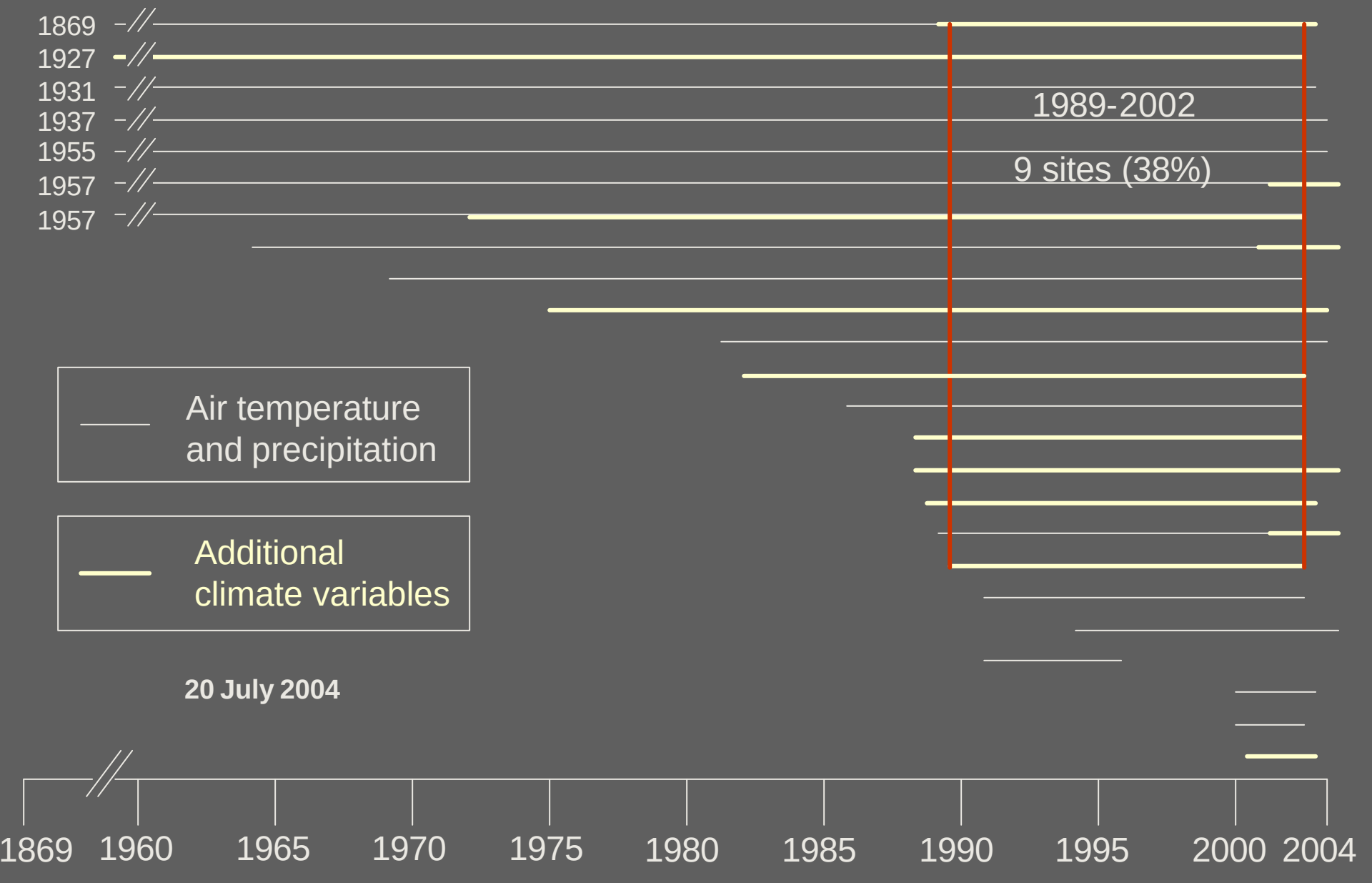
ClimDB Temporal Coverage By Site

Air temperature and precipitation



ClimDB Temporal Coverage By Site

Additional climate variables (Solar, wind, rel. humidity, soil temp., ...)



Developments in Wireless Sensor Networks

Barbara Benson

North Temperate Lakes LTER

The Current Model

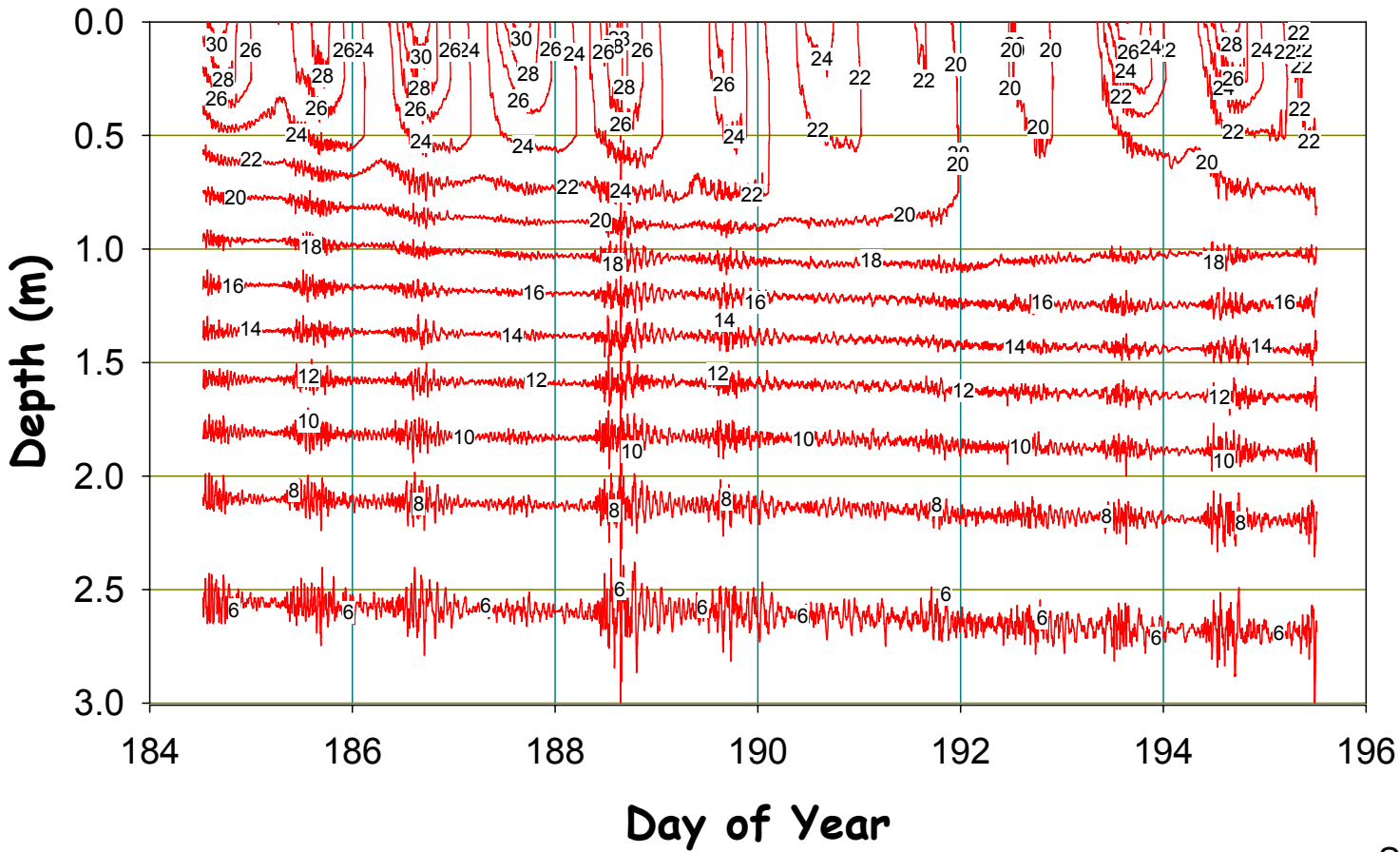
Instrumented Platforms

- make high frequency observations of key variables
- send data to web-accessible database in near real time

Source: Kratz

Sensors Produce Richness in Data
Allow researchers to observe the unobservable

Trout Bog Buoy Data 2003
Water Temperature



Source: Kratz

Some Specific Scientific Drivers

at the interface of biological and physical limnology

- How do lakes influence atmospheric carbon and water budgets through atmospheric exchanges at local, regional and global scales? (Lake Physics)

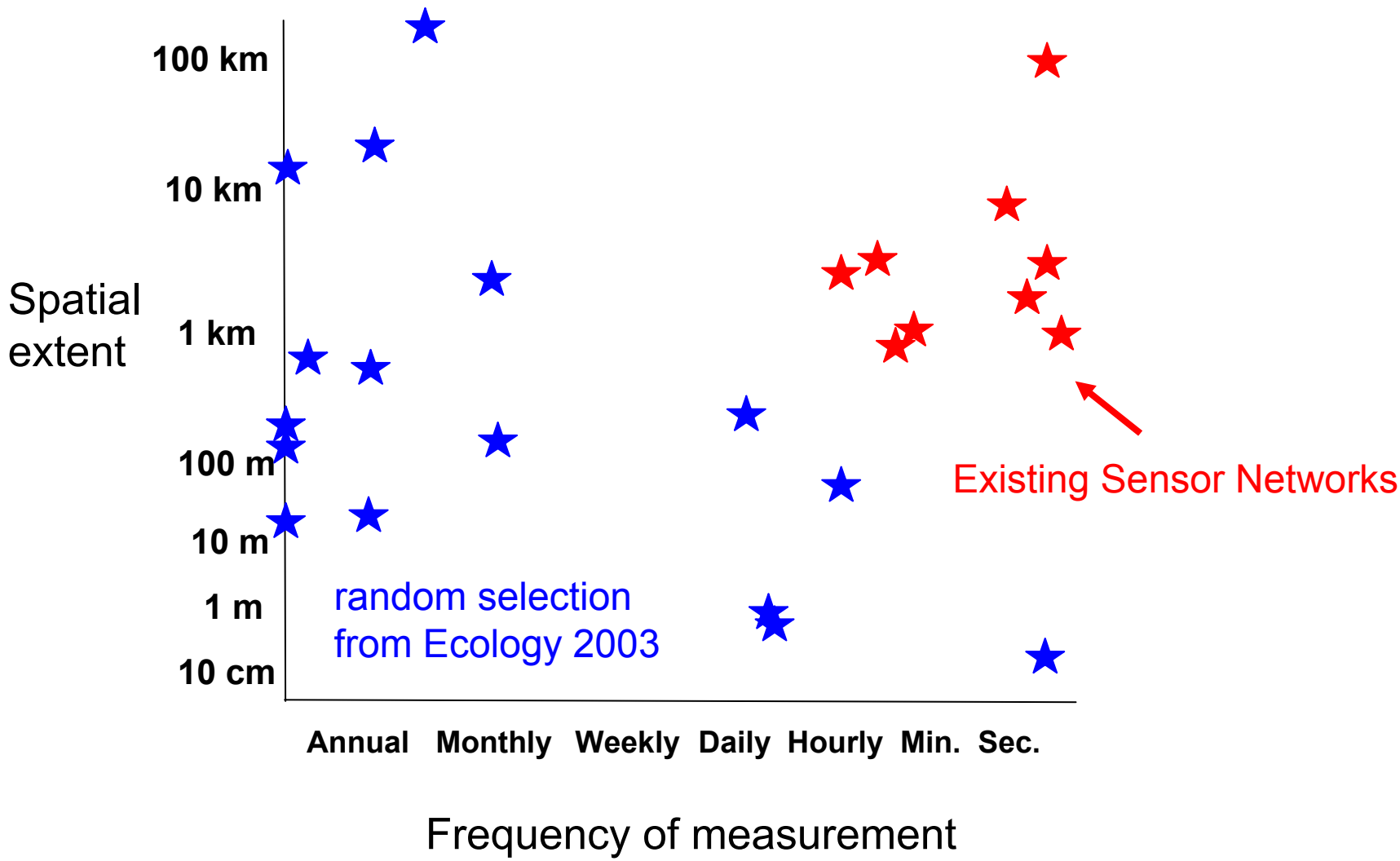
- What are the long-term effects of physical disturbances on the biology and chemistry of lakes? (Disturbances)

- How do nutrient loading, hydrology, geologic setting, and climate regime influence the metabolic balance in lakes? (Lake Metabolism)

Questions at Different Temporal Scales: at interface of physical and biological limnology

- Seasonal:
  - Under what conditions are lakes net sources of carbon to the atmosphere?
- Weekly:
  - How does lake metabolism respond to episodic driving events, such as thunderstorms or typhoons?
- Minute or hourly:
  - Under what conditions do physical, rather than biological, processes drive observations of dissolved oxygen dynamics?

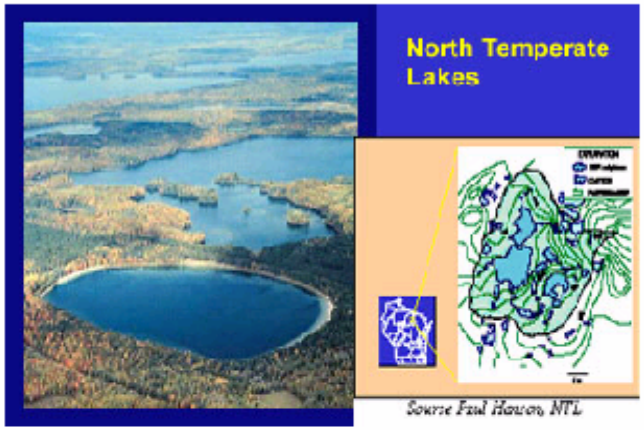
Sensor networks allow high frequency observations over broad spatial extents

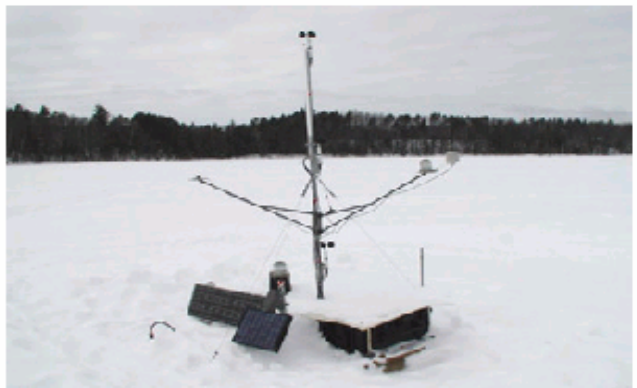


Source: John Porter et al., Bioscience, Vol 55, pp 561-572 July 2005

中文版

Collaborative Lake Metabolism Project

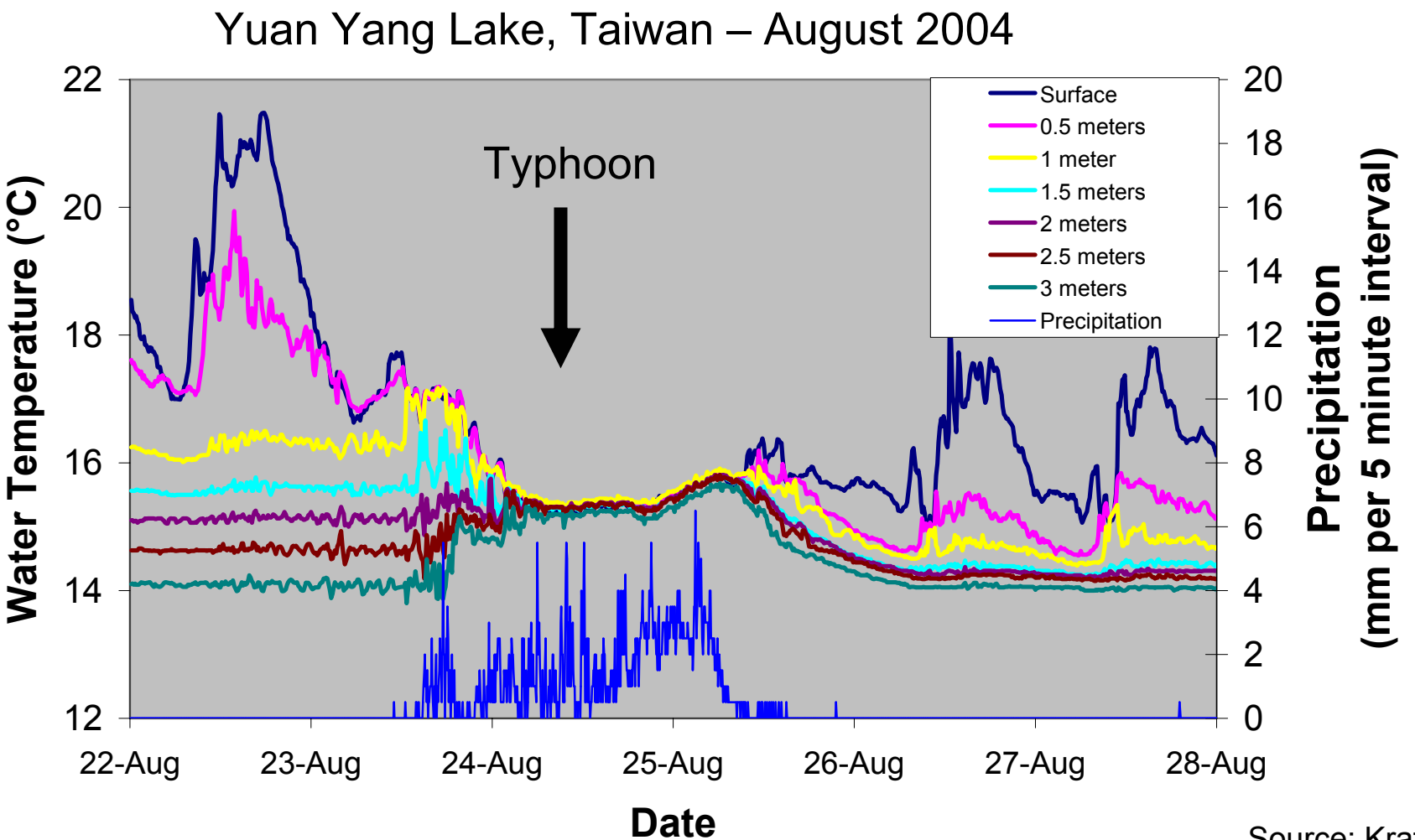




http://lakemetabolism.org

Photo by Peter Arzberger

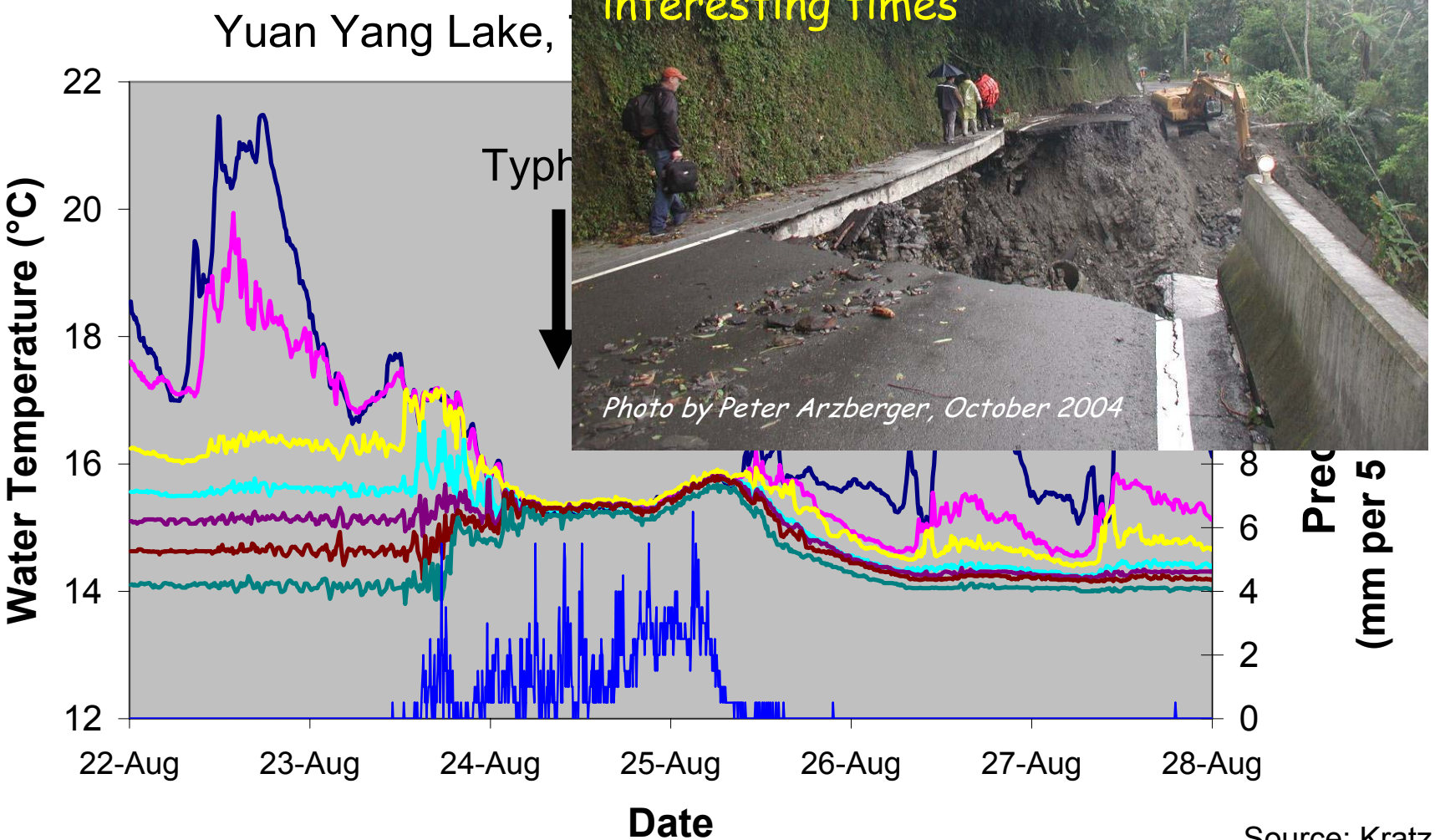
An example of episodic events and threshold dynamics



Source: Kratz

Part of a growing global lake observatory network - http://lakemetabolism.org

An example o
thres

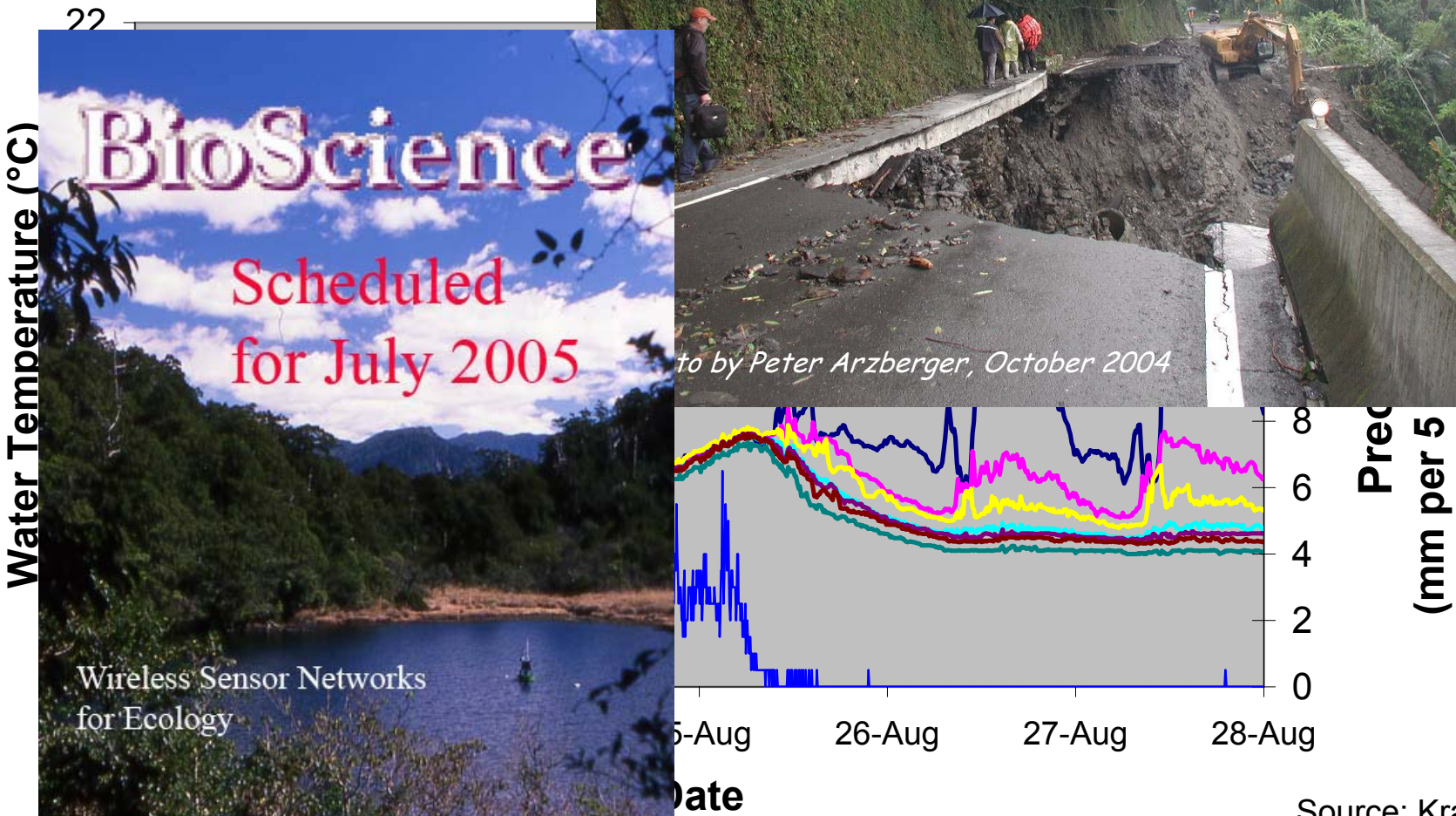


Source: Kratz

Part of a growing global lake observatory network - http://lakemetabolism.org

An example o
thres

Yuan Yang Lake,



Source: Kratz

Part of a growing global lake observatory network - http://lakemetabolism.org

Lake Metabolism Project --- Data

Welcome: bjber
Mar

Aug 1, 200

LAKE DATABASE : YUAN YANG LAKE DATABASE

SAMPLE RATE : DAILY

VARIABLE SET : YYL_BOG_BUOY_DAILY --> AVG_GPN_DOSAT & AVG_GPN_WTEMP

DATE RANGE : 2004-06-01 ~ 2005-07-31

DATA RETRIEVAL : Display in Web Browser Download ( )

SQL STATEMENT :
(generated for data retrieval)

select sampledate, AVG_GPN_DOSAT, FLAG_AVG_GPN_DOSAT, AVG_GPN_WTEMP, FLAG_AVG_GPN_WTEMP from YYL.YYL_BOG_
where sampledate >= TO_DATE('2004-06-01', 'YYYY-MM-DD') and sampledate <= TO_DATE('2005-07-31', 'YYYY-MM-DD') order
sampledate

Toward a distributed information system for marine biology and limnology

Co-Host meeting to consider global lake and coral reef observing network March 2005

Source: Arzberger

Gordon and Betty MOORE FOUNDATION

Building the Community

- Science Questions
- Architectural Design
- Broaden Involvement at all levels

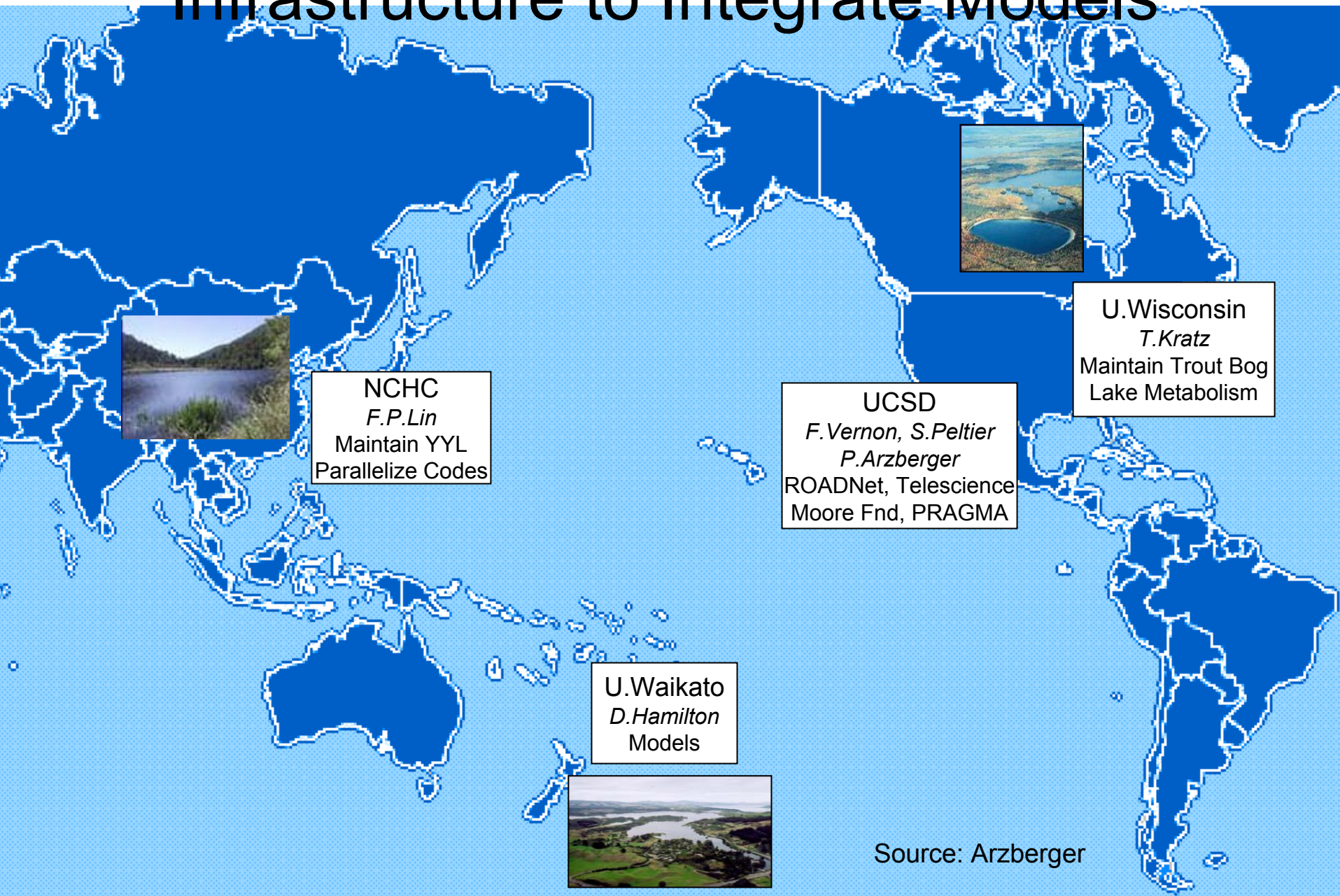
GLEON

global lake ecological observatory network

A foundation for GLEON

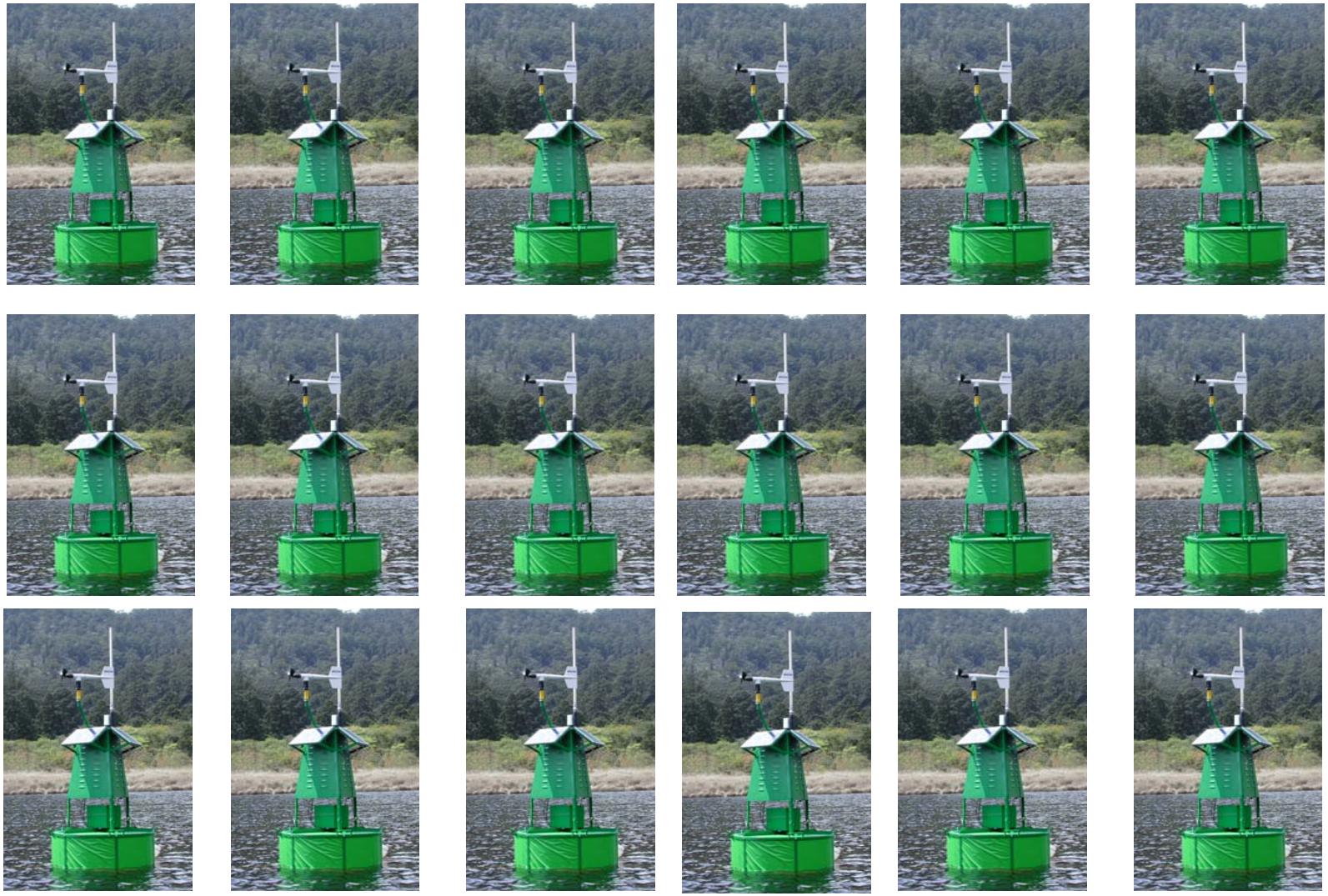
Source: Arzberger

Team e-Science: Building End-to-End Infrastructure to Integrate Models



Source: Arzberger

Scalable instrumentation and cyberfrastructure is critical



lakemetabolism.org

Source: Tim Kratz

Problematic, but possible with today's cyberinfrastructure

Scale needed to answer regional/continental questions

Addressing the Scaling Challenge NSF NEON Award

- Collaborative Research: Automating Scaling and Data Processing in a Network of Sensors: Towards a Global Network for Lake Metabolism Research and Education
  - UCSD, UWI, IU, SUNY-Binghamton
- Automate
  - Configuration of Sensor
  - QA/QC and Event Detection
- Service Oriented Architecture
- Leverage Activities
  - ROADNet
  - NMI

Source: Arzberger

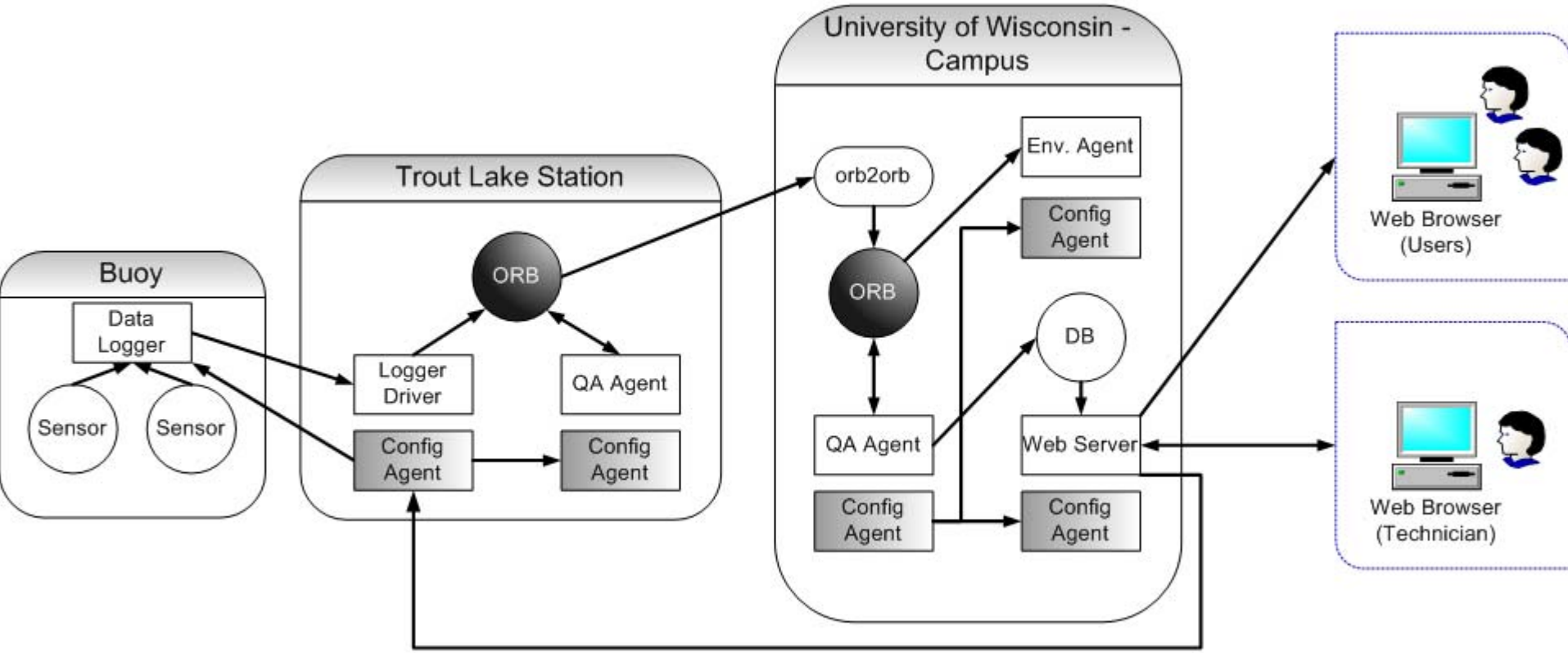
Major goals

- Develop new methods and tools to help automate the updating of data flows from dynamically deployed sensors to publicly accessible biological databases
- Develop a suite of new algorithms and software for analysis of biological information to automate detection (real-time) of events based on data from sensors allowing rapid response

Automated Scaling and Data Processing in a Network of Sensors

Collaboration with computer scientists at UCSD, Indiana, and SUNY-Binghamton and computer engineering colleague at UW-Madison

Growing the Sensor Network: Agents and Remote Code Deployment



NSF NEON Grant: Addressing the Scaling Challenge

Advanced tools for automated and intelligent processing and control

- Automated QA
- Machine learning for event detection
- Real-time intelligent agents
  - Determining and executing appropriate response to detected events
  - Based on statistical decision theory

Dissemination of results to the broader research community

- Web page

  lakemetabolism.org
- Special workshops
- Annual PRAGMA brochure

Cyberinfrastructure Planning within the LTER Network Planning Grant Context

Barbara Benson
James Brunt
Peter McCartney
John Vande Castle

Preparing the LTER Network for Collaborative Science, Education and Synthesis: A Planning Proposal

Jim Gosz

Barbara Benson

Dan Childers

Scott Collins

Allison Whitmer

Objectives

- Develop a plan for LTER network-level science, technology, and training
- Explore alternative governance, planning and evaluation structures for managing LTER Network science
- Envision and plan for education, training, outreach, and knowledge exchange activities to link LTER science with application needs

network level science

- New initiatives in long-term thematic, regional, and network-scale science
- Increasing the capabilities of scientists/sites (e.g., cyberinfrastructure, technical expertise) to perform research and education for the new environmental challenges

The LTER Network will identify and pursue the appropriate strategy to accelerate its transition from an association of sites driven by local goals and resources to a fully functional Network driven by regional and national research priorities and shared resources.

Grand Challenge Conceptual Domains

- Alterations in biotic structure/biodiversity
- Altered biogeochemical cycles at multiple spatial scales
- Climate change and climatic variability
- Coupled human-natural ecosystems

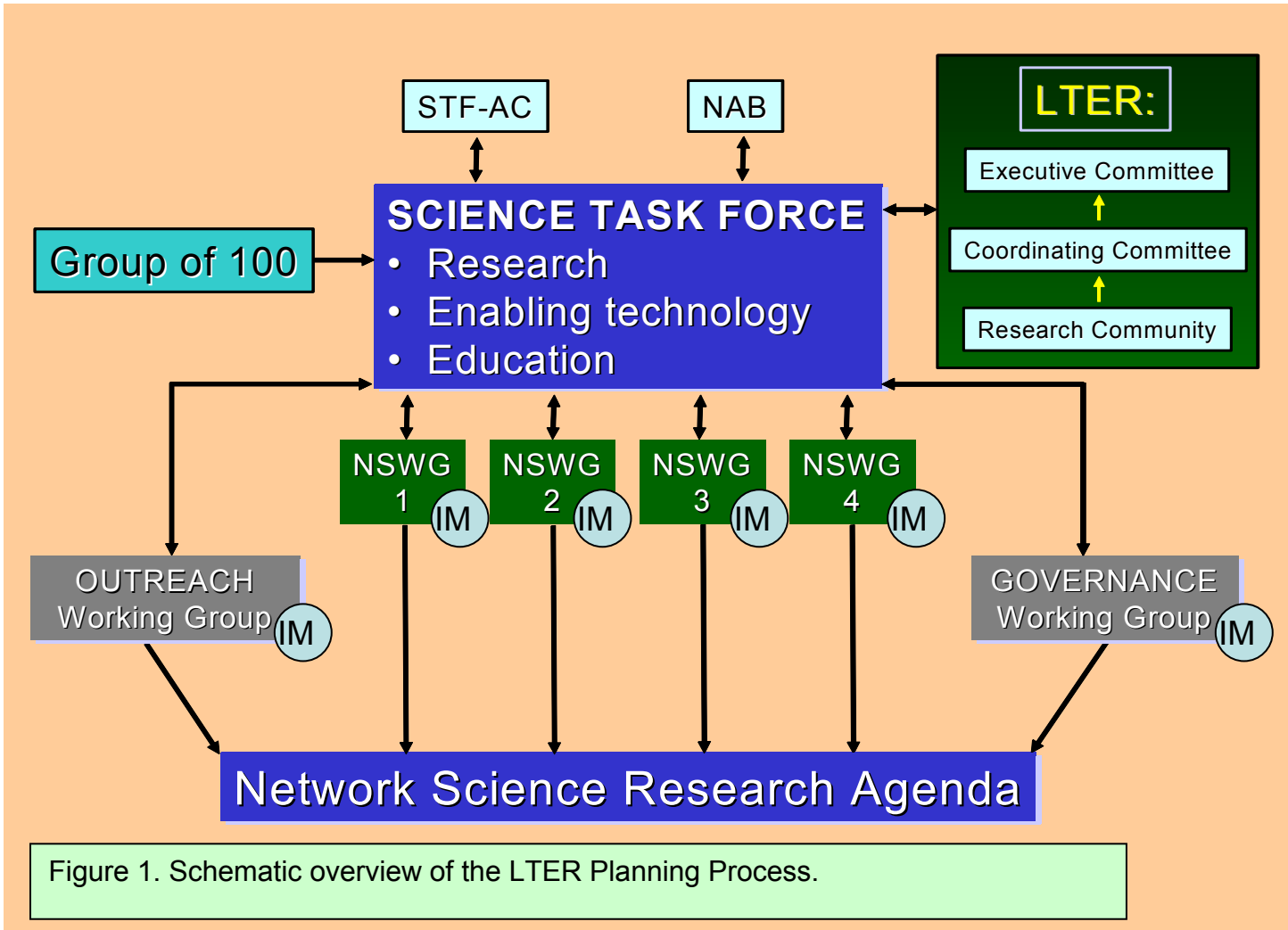


Figure 1. Schematic overview of the LTER Planning Process.

IM Participants in Working Groups

- Biotic structure: Corinna Gries
- Biogeochemistry: Kristin Vanderbilt
- Climate: Don Henshaw
- Human Dimensions: Jonathan Walsh
- Education: Ken Ramsey
- Governance: Karen Baker

By the end of the 24 month process, the STF will synthesize a network research agenda that will serve as the basis for developing at least one integrated research and education proposal to the NSF LTER Program to implement our new, synthetic, network-level long-term ecological research and training activities.

Report from IM Participants
June 2005
Santa Fe meeting

Cross-site Science Activities and Cyberinfrastructure (CI) needs

- Synthesis of existing data and new cross-site experiments
- Tools for collection and management of large data volumes
- New tools for data analysis
- Enhanced model management
- Collaborative communications technology
- Special data needs of education, outreach, and training

Data Synthesis

- Availability of LTER and non-LTER data and structured metadata
- Search and access capability of LTER site data through a single web portal
- Mechanisms for mining data from LTER and non-LTER sites: discovery, access, and integration of data
- Harmonization of different units, different temporal and spatial scales
  - Development of ontologies/semantic mediation system
  - Development of controlled vocabularies (units and measurements)

Collection and management of large volumes of data

- Wireless sensor networks
- Genomic array data
- Disk space and network transmission needs will be high
- Quality assurance tools for real-time data
- Tools for automated generation of metadata

Spatial data

- Coordinated acquisition of data
  - Remote sensing imagery (tasking satellites)
  - Demographic data
- Infrastructure for storage and management of large quantities of spatial data will be needed
- Integration of spatial and “non-spatial” data

Management and analysis of data from cross-site experiments

- Development of standardized measurement protocols
- Assembly of cross-site data sets into centralized, web-accessible databases
- Tools for analyzing data collectively: web site could provide query tools, statistical summaries, visualization tools
- Quality assurance
- Metadata generation

Analytical tools

- Tools that facilitate data synthesis that allow the definition of complex workflows involving heterogeneous applications, models and data
  - Automate metadata generation of derived products (e.g., lineage)
  - Automate and document workflows

Models/Model Management

- Need for shared community models
  - Forecasting, exploratory analysis, simulation
  - Available for scientist use with GUI interfaces
  - Can be applied across sites
  - Available source code
  - Better documentation
- Consistent model metadata representation
  - Basis for the model, sensitivity and uncertainty descriptions, parameter sources, input/output descriptions, and model validation
- Versioning system for models needed
- System for archiving model output needed
- Coupling of models, e.g., associate physical and biotic, inter-scale from genome to ecosystem

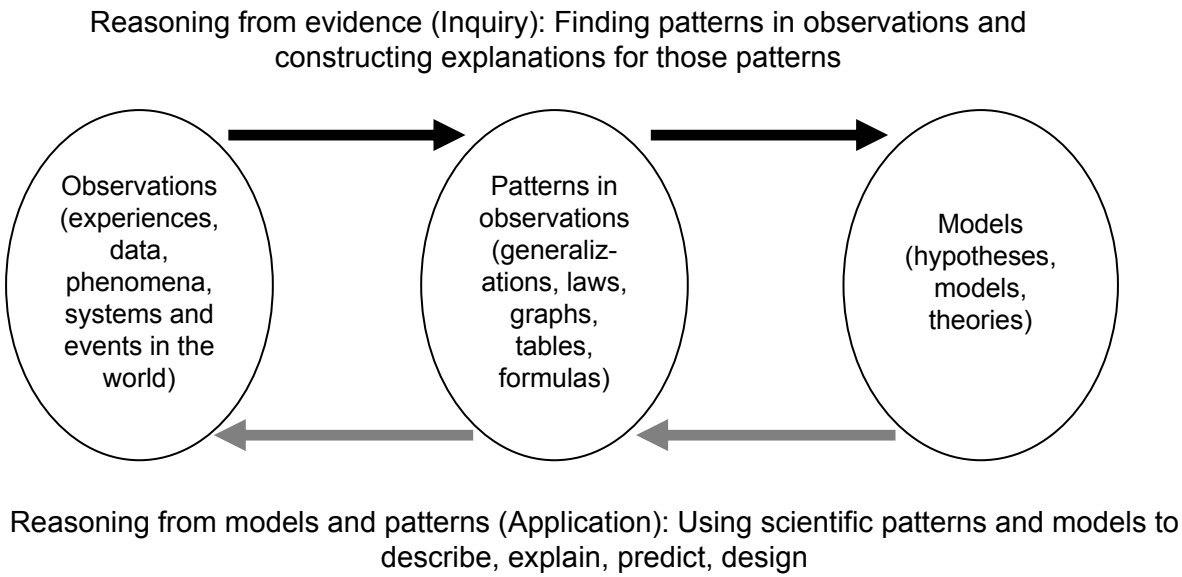
Collaborative Tools

- Investigators at different sites will use access grid, polycoms, wiki sites, user forums, document sharing, other technologies for communication.

Acquisition or Collection and Management of Socioeconomic Data

- Issues:
  - Research on human subjects requires approval by Review Boards
  - Privacy/anonymity of respondents
  - Data sharing agreements may be required
  - Research may require interactions with and relevance to subjects
- Opportunities:
  - Publicly available data sets are often collected regularly over time, for entire country, and allow researchers to detect trends in social and economic systems
  - Data management protocols are identical for those of biophysical data

Education Data Needs



- Databases and applications to support sharing curriculum and training materials
- Develop contact database of organizations, agencies, institutions of higher education that develop and/or impact state/national education standards

Additional Bullets

- Qualitative data collection technologies
- Tools that support easy access to citations of data sources and compliance with data policies (acknowledgement, etc.)

CI Supplement

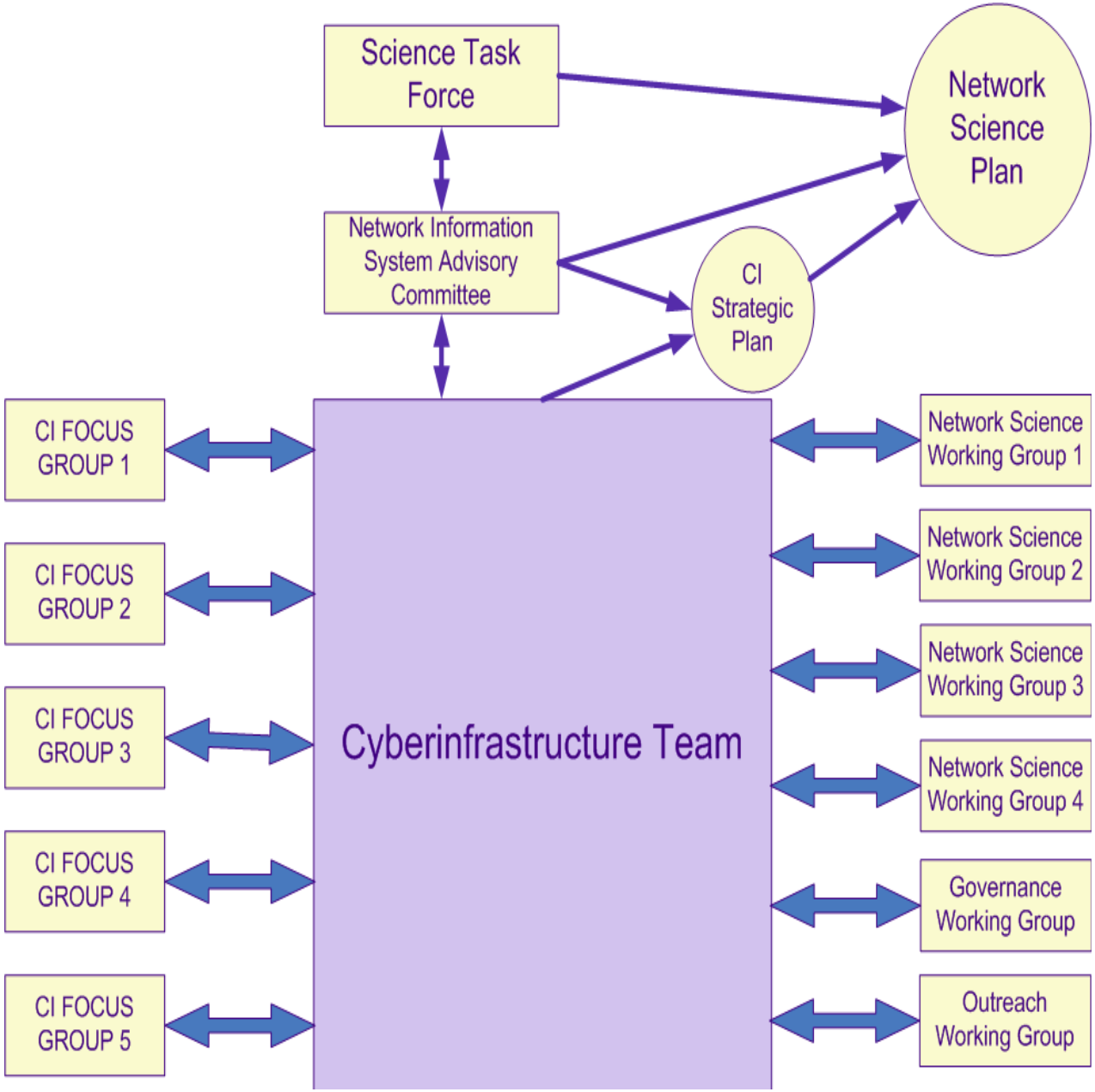
- strengthen the LTER planning effort through a broadly based consideration of needed cyberinfrastructure
- engage computer and information scientists to address the new integrative challenges presented by the expanding spatial, temporal and interdisciplinary scope of LTER network science
- provide cross-fertilization between LTER CI planning and that of other concurrent efforts within and beyond the ecological science community.

Charges to the CI Working Group

- identify potential areas where CI is required to support science emerging from the LTER planning process
- assess the CI capabilities of LTER and non-LTER resources
- organize focused team meetings around major areas of identified CI needs
- develop a design and implementation strategy to provide the necessary CI

Themes: CI Supplement

- *Sustainability and human capacity*
  - Training, tools, staff for implementation and maintenance
- *Applications*
  - Model integration systems, complex workflows
- *Data Acquisition*
  - Data ingestion and integration; sensor networks, remote sensing
- *Distributed Computing*
  - Communication and collaboration technology; grid computing and resource sharing
- *Knowledge Representation*
  - Document semantics, units, methods of measurement


Science Task Force
Network Science Plan
Network Information System Advisory Committee
CI Strategic Plan
CI FOCUS GROUP 1
CI FOCUS GROUP 2
CI FOCUS GROUP 3
CI FOCUS GROUP 4
CI FOCUS GROUP 5
Cyberinfrastructure Team
Network Science Working Group 1
Network Science Working Group 2
Network Science Working Group 3
Network Science Working Group 4
Governance Working Group
Outreach Working Group

nonLTER CI Team members

- Chaitan Baru (GEON, NEON)
- Bryan Beecher (ICPSR)
- Bob Cook (DAAC)
- Peter Cornillon (OPeNDAP)
- Mandy Lane (ALTER-Net)
- Mark Schildhauer (NCEAS)
- Mark Stromberg (OBFS)
- Michael Piasecki (CUAHSI)
- Mike Frame (NBII)

Timetable and milestones

June 16-17,2005 CI Team +NISAC +Embedded IM reps

Create outline for CI strategic plan; develop working materials, assign topics, and assemble review documents for CI Focus groups; and create instructions for CI needs assessment by NSWGs

July – Dec 2005 four CI Focus Groups

Perform review, needs assessment and develop whitepapers on assigned CI topics

Sept 2005 CI Team + STF Briefing with NSF program officers

Nov 2005 Network Science Planning Meeting

Jan 2006 CI Team + Embedded IM

Integrate Focus Group whitepapers and NSWG needs assessments into final CI strategic plan

Mar 2006 CI Team +NISAC

Review CI plan with NISAC and finalize for STF

May 2006 NISAC + STF

Merge CI strategic plan into overall network plan

8 Sept 2006 LTER All Scientists Meeting with partners

Roll out of network science and CI plan and implementation workshops

Products

- Strategic plan for cyberinfrastructure based on
  - Technical whitepapers
  - Science needs assessment

- Cyberinfrastructure proposals

Focus Groups

- Infrastructure for data integration
- Modeling frameworks
- Framework for x-site/Network experiments

- Systems architecture & human capacity & funding strategies

Use Case Scenarios

to ground the cyberinfrastructure planning in realistic and complete visions of the scientific needs

- What is the science/teaching question?
- Describe in detail an end-to-end narrative for the workflow.
- What are the limitations?

Describe in detail an end-to-end narrative for the workflow

1. data needed – new data collection, data mining, existing monitoring
2. quality control, data integration issues
3. who is involved
4. analytic and modeling tools needed
5. network, computational needs

What are the limitations?

1. cost
2. lack of resources, ability, expertise
3. replicability, auditability
4. need for new technology
5. need for new organizational structure for collaboration
6. others?

Needs Assessment

- Assembled and distributed by John Vande Castle

- https://www.surveymonkey.com/s.asp?u=935391203229

Agenda

- expectations & goals
- what do we want ontologies to do for us?
- are there common measurements within lter?
- what information is needed to describe an attribute?
- Implementation issues
- common framework
- user interfaces both for query and submisssion
- EML integration

What is an ontology

- set of related concepts or classes
- classes can be inherited from other classes
- classes can have properties
- Ontologies can be very deep or very shallow
- Tools exist to traverse and query ontologies
  - eg. find sibling concepts of a class through their common parent class

Building ontologies

- Top down – start with very abstract concepts and fill in detail
- Bottom up – define lowest-order concepts, then start looking for grouping concepts.
- Middle out – define reasonably low-order concepts but avoid extreme detail introduced terms that merely mix concepts

What goes in our ontology?

- To fully evaluate, use, or rescale an ecological observation one must understand its properties:
  - Measurement type
  - Context: Space, time, culture, deposition, substrate
  - Methods
  - Units
  - Accuracy

Measurement Type

- formally register the concept for what is being observed (could be classifications, measurements, presence/absence)
  - ideally divorced from other properties such as context methods, units (contrasts with EPA Storet which defines a unique concept for each combination of measuremnt, units, methods)
  - measurement scale& domain information
    - however, domain within given instance may be further restricted by context – body length vs body length of voles

What do we expect from this

- simplify or reduce work involved in filling out eml-attribute
  - reusing information, eliminating redundancy
  - enabling templates in metadata creation
- Improving applications
  - using attribute ontologies as indexing and keywords
  - linking attribute ontologies to display rules
- Support for data integration
  - linking attributes to rules for data aggregation, reclassification, calculations

Semantic Issues

- Dealing with attribute descriptions that embedd some contextual constraint
  - eg soil temp, air temp, min temp are all conextualized observations of temp
  - virtually all count and proportion based measures are context laden – taxonomy or chemistry
- not clear rules yet
  - guided by practical limits (taxonomic abundance)
  - what resonates with scientists (chemistry)

Pilot pilot-studies

- water chemistry
  - found common variables eg DOC.
  - found many qualified by context (GW, stream)
- biodiversity
  - konza datasets five biomass columns that were subclasses of live biomass restricted by woody, annuals, etc.
  - sampling was an important but difficult concept to evaluate.
- ANPP

SiteDB Working Group 1, 2005 Annual IM Meeting

We explored the utility of siteDB to deliver information, metadata, and data from across LTER sites in a uniform way from a centralized portal. We divided our discussions into the design and navigation to and from siteDB and the expansion of content in siteDB. Links to siteDB from the LTER Network homepage should be easier to find. We suggested ways to improve navigation to siteDB, network resources and sites' websites by naming each type of link more descriptively and consistently in each place it appears. We would like to see links to the sites' websites emphasized and maintained at the forefront of siteDB. Specific links from siteDB to categorical web pages on a sites' website should navigate to similar information from website to website. Recommendations regarding the expected content under suggested links on website homepages may address these inconsistencies. We also discussed ways to expand siteDB. We need improved ways to insert and update content as to not copy and paste information redundantly from within a site's website to siteDB. We can improve delivery of content from siteDB by creating more queries to group and summarize information. We discussed site descriptions to complete and/or supplement next, which may include climate, geology and societal characteristics. Discovery of data may be centralized by installing the metacat query interface in siteDB. Other kinds of data may be accessed through siteDB, such as remotely sensed data, and a list of long-term core data sets. The working group will continue to discuss these issues over the next year.

Web Site Recommendations Working Group 4, 2005 Annual IM Meeting

Our working group discussed the development of recommendations that are referenced in the newly adopted IM review Criteria. We will continue to document these recommendations and then they will need to go through an approval process within the greater LTER community. The recommendations may be applied to new and re-designed sites' websites to improve the delivery of information and to help portray them as part of the LTER Network. We would like to provide examples and descriptions of good website design practices within the recommendations. We focused on what information should be linked directly from the homepage. We discussed a set of consistent links, menu or other navigation tools for homepages. We gave examples of descriptive versus ambiguous names for the links, and discussed the type of content an end user would expect to find under each link. The navigation menu on the homepage should be repeated on each page of the web site as to facilitate finding information. Scripts and tools with drop down menus that link to other LTER sites and LTER Network resources (i.e. climDB, siteDB) should be installed on the homepage, as well as a statement such as "this site is a member of the LTER Network". These scripts and text may be developed and distributed by the Network office. Finally, we reviewed proposed designs for the LTER Network website and provided feedback to web designers.

LTER Information Manager Meeting 2005

Working Group 7 Report (KBaker, BBenson, MJones, FMillerand)
Community Process and Standards Implementation
August 5, 2005 (5:30-6:30pm) Montreal

The Community Process Working Group provided a forum for accumulating experiences with EML and a distillation from these experiences of some principles and critical questions. The event together with a conceptual diagram and surveys provided an opportunity to consider the process of standards' development.

Some lessons learned were defined at 3 levels:
(1) The Design and Development process: a shorter cycle would allow time for testing; the role of early-adopters may be extended so that they may act not only as testers but they also may make an important bridge with the community; funding is needed to support such efforts.

(2) Deployment and Enactment process: the alignment of timing and expectations i.e. the need to align the levels of 'readiness' at each site with regard to implementation level expectations; need for a site representative to act as a liaison with the network.

(3) Maintenance and sustainability of community standards: The IM community has invested in and contributed a lot to EML. This leads to how to bring the IM community into discussions about the future of EML? Who is going to do the support for EML (no more funds at the developers)? Could this be a question for NSF?

The concept of 'enactment' from social informatics research was proposed as a way to augment understanding of a standard implementation process. It highlights the critical step of local implementation that necessarily implies an adaptation of local practices to the standard and the modification of the standard itself. The standardization process was represented in a cycle of 4 major processes: design, development, community level deployment, and local enactment.

Site Information System Innovations Working Group

Reporter: John Porter

A small working group met to trade information on recent innovations in information management at different LTER sites. The group did "round-robin" discussion with each participant identifying their most useful innovation, and answering questions on how it was used and/or implemented. Specific topics were:

- **Adding full-text electronic versions of LTER publications to a site's web page.** The discussion here focused on legal issues, with the lawyers at some institutions noting the pressure by NSF and others on publishers to allow posting of federally funded research results on web sites of the institutions receiving the grants. Analogies with the fair-use distribution of reprints by authors (often in electronic form) were discussed. Other sites had received legal advice that they could not post such documents under any circumstances. Although the legal issues are cloudy, there was a unanimous sense that such content would be widely desired by LTER researchers.
- **Moving to dynamically generated web pages**. Many LTER sites already implement at least some content dynamically, where a database is used to populate a web page "template" with the desired information. Here we discussed both customized systems and generalized "content management systems" (such as PostNuke).
- **Integrating LNO components into site web pages**. Some sites are now using databases maintained by the sites at the LNO, such as personnel and publications databases, to replace local versions. Abilities to restrict searches to particular sites or to extract data (but not the entire web page) from the LNO were discussed, along with the need for more clearly documented interfaces and APIs.
- **Simple tools for the web display of sequential images.** A simple JavaScript application can display sequentially numbered .jpg or .gif files.
- **Progress on tools to aid in EML production.** Under development are tools for mapping meta-database schemas to EML and work on harvest lists.
- **Display of "preview" data subsets to accompany metadata.** For users, often a "picture is worth a thousand words" – display of the first 10-20 lines of a dataset can help users interpret the metadata.

Despite the limited number of participants, we found the information exchange to be very valuable. We hope that alternative activities, such as having each LTER site provide a brief description of the "three things" that they are most proud of, or receive the most enthusiastic response by users, along with a link to the pages, so that other IMs can "browse" for features to add to their own systems.

Design of Query Interface to LTER Catalog: Controlled Vocabularies Working Group

Group members: John Porter (reporter), Jonathan Walsh, Suzanne Remillard, Jingcheng Gao, Inigo San Gil, Duane Costa and Tim Rhyne

This working group built on discussions at the 2004 Information Managers Meeting regarding the desirability of creating "browsable" data and improving data discovery through searching. The group developed a strategy for evaluating existing resources and possible collaborations in the development of a hierarchical controlled vocabulary for describing data sets.

The group began with a general discussion of objectives and specific needs, with reference to existing systems. For example, NASA's "Global Change Master Directory" has developed a hierarchical system of keywords and key terms which they characterize as "Category", "Topic", "Theme", "Variable" and "Keyword". The four highest levels of the hierarchy are based a controlled vocabulary of specific terms. The lowest level (keyword) allows researchers to pick their own terms. This hierarchical system allows a "browsable" page where researchers burrow down from the most general terms to the most specific (http://globalchange.nasa.gov/). The National Biological Information Infrastructure (NBII) uses a more general thesaurus, including broader and narrower terms to help facilitate searching its data.

Currently the keywords used to characterize most LTER datasets are uncontrolled, meaning that they are specified entirely by the data creator. One of the challenges facing LTER and external researchers in discovering data from LTER sites is inconsistent application of keywords. A researcher interested in carbon dioxide measurements must search on "Carbon Dioxide," "Carbon Dioxide Concentration", "Carbon Dioxide Emission", "Carbon Dioxide Evolution", "Carbon Dioxide Flux", "Carbon Dynamics", and "CO2" (all from the August 2005 Data Table of Contents). Moreover, additional key terms such as "gas flux" coupled with "carbon" may indicate data of interest. A study of keywords several years ago indicated that the modal number of LTER sites the typical keyword was used at was 2! The LTER MetaCat system does implement a rudimentary browse interface with six major categories with 15-20 terms each (total ~80), but the concensus of the group was that a richer hierarchy would be valuable addition.

The working group started with a brainstorming activity to identify possible ways of improving data discovery through the use of controlled vocabularies, with the goal of improving searchability and improving the "browse" interface. The "browse" interfaces have the advantage over search in that researchers seeking data can move down a hierarchy from the most general to the most specific, with no "null" searches, because each term in the browse tree has at least one dataset associated with it or it would not be displayed. Additional uses of structured controlled vocabularies would be to enhance searching by facilitating display or searching of related terms.

The group identified several sources of controlled vocabularies or thesauri that are already in use or under development by environmentally-related projects. These included:

- NBII Thesaurus (http://nbii-thesaurus.ornl.gov) is under development, with a draft web services interface already available using a Simple Knowledge Organization System (SKOS) interface.
- The General Multilingual Environmental Thesaurus (GEMET) is under development. Resource Discovery Framework (RDF) files are available and there is an associated web service.
- The Global Change Master Directory has a hierarchical set of key terms used to index its data at: http://gcmd.nasa.gov/Resources/valids/gcmd_parameters.html
- The Integrated Taxonmic Information System (IT IS) has taxonomically-related keywords. A web-services version called SPIRE is available from the University of Maryland, Baltimore County.
- There are numerous "gazetteers" that relate spatial coordinates with names.

It also identified sources of existing (albeit, largely uncontrolled) sources of keywords derived from LTER datasets. These included:

- Tagged fields in EML metadata documents including:
    - Title
    - Keywords
    - Attribute definitions
    - Taxonomy
- Controlled keyword lists from individual LTER sites, such as the the H.J. Andrews LTER hierarchical controlled vocabulary.
- Keywords from the LTER Data Table-of-Contents (DTOC)
- Words from the title or keywords sections of published documents produced by the LTER network.

The working group strategized that the best approach would be to evaluate the existing controlled vocabularies' suitability for use by LTER. The criteria for suitability discussed were:

1. Is there a broad overlap with existing LTER keywords? A broad overlap with existing LTER terms will facilitate implementation of the controlled vocabulary and reduce the number of terms that need to be added. .
2. What additional capabilities will the controlled vocabularies provide? Ideally, the controlled vocabulary should be structured so that more general or specific terms can be extracted and that synonyms and related terms would also be found during a search.
3. What is the availability of the controlled vocabulary, and are there ways LTER could participate in its alteration if specific needs are identified?
4. Would the controlled vocabulary help support searches across language barriers so that US LTER data would be available to a wider array of international scientists? If adopted by other international projects, controlled vocabularies would also facilitate access by US scientists.

The working group proposed a process for evaluating the suitability of use for the different controlled vocabularies.

1. Accumulate a master list of existing LTER keywords from the sources listed above.
2. Identify broadly vs. narrowly applied LTER keywords by tallying the number of datasets or sites using specific keywords.
3. Search candidate controlled vocabularies for those terms and tally the number of matches.
4. When there is a match, evaluate the "richness" of the information returned by tallying the number of related, broader or more specific terms found.
5. Investigate the availability and openness of different controlled vocabulary maintainers to LTER participation in updating and expansion of the controlled vocabulary.

If one or more suitable controlled vocabularies are identified, the next step is to develop prototype applications such as tools to "enrich" existing EML documents by adding controlled terms, search tools that query related keywords that may not actually occur in the document and tools that create a "browse" interface based on hierarchical controlled vocabularies. These prototype applications could be evaluated by researchers, and if found useful, production implementations pursued.

The working group also identified individuals willing to take on specific tasks including developing lists of keywords from different sources (Costa, San Gil, Porter, Gao and Remillard), testing against existing thesauri (Walsh, Rhyne, Bohm, Porter and Gao) and developing prototype applications (Costa and San Gil, subject to NISAC approval).

Emphasis was also placed on communicating results of these activities to the LTER IM group and to the broader community through periodic communications and, ultimately, publication.

All Scientists' Meeting 2006 Planning

Summary of Working Group Session

- September 19-24 dates blocked out.
- Six sessions three days
- Estes Park good place to get work done, good for product oriented content
- CI planning and implementation (a transfer of information) as it comes out of the Network planning activities, other ecoinformatics community members contribute here to proposals and products for the next steps

- IM talks last ASM were mostly attended by other IMs, and it was hard for IMs to attend other talks.
- Don't make too many sessions
- Internation folks nice but, expensive.
    - Need to start early
    - Might have 75k… That only brings 35 people from other countries and there's lots of non-IM people coming so we wont have much luck getting all the money we need.
    - Better determine our needs now then ask Henry for some more money.
- Presentation and q and a format didn't work out so well with regard to info mgt sessions
- Roundtable?
- Plenary?
- Plenary session for IM or embed an IM in a plenary session?
- May not be a plenary session – they didn't review well last year
- IM plenary last time was rejected by planning committee
- If we could get someone compelling and interesting

- Preliminary List of Topics
    - Advance training for NIS strategic plan
    - CI planning and implementation
    - Sensor network workshop
    - Interoperability @ national level so field station can exchange data seamlessly at national level. Across languages and alphabet. Avinash has something to say and good information about that
    - Bridging data discovery across languages and alphabets
    - Bridging XML across languages
    - Tools to foster networks and networks of networks
    - Something co-organized between US IMs and international IMs.
    - Some sort of GIS workshop – common ways of overcoming problems like EML or integrating site science

    - Cyberinfrastructure perspective – have some scientists actually do some presenting on CI. Good way to engage scientists. Could tap graduate students same way too.
    - Ontology
    - Unit dictionary
    - Observing the previously unobservable
    - Workflows
    - Computational Grid
    - Social Science data
    - S. Africa
    - Interoperability
    - Languages

- Short List of Topics
    - Advance training for NIS strategic plan (Henshaw)
    - CI planning/implementation
    - IM, NISAC, CI Team
    - Sensor network workshop
    - GIS/remote sensing --- Teresa Valentine
    - Ontologies
    - Orange box
    - Observing the previously unobservable
    - Estrin, Gage, NEON
    - Self-establishing networks, tiny OS
    - Workflows: Kepler, D2K, I2KComputational Grid perspective Chetan Baru
    - Managing sociological data
    - Organizing w biogeophysical scientists
    - Roberta, head of Columbia Univ. DAT, Gragson
    - Legal issues

LTER Annual IM Meeting, Montreal, Canada, Friday Aug 5, 2005
Working Group Session I, Group 1: Web Access to Information: SiteDB

This working group was formed to discuss ways to enhance access to information about LTER sites online through SiteDB. The objective of siteDB, the LTER Site Description Directory module is to make available basic information about site characteristics for the LTER network of sites. General site descriptive information includes specifics such as history, classification type, latitude, longitude, area, elevation, and site contacts (http://pal.lternet.edu/lter/dm/projects/sitedb/overview.html). SiteDB may also serve as a web portal to other information and data from sites. Eda Meléndez-Colom reported current participation in siteDB and this information will be made available as a reference for the meeting. Now, we would like to take this opportunity to discuss and identify what aspects of siteDB should be improved upon or amended. This working group was formed to explore the quality and quantity of information, which may accessed through siteDB as a centralized, uniform portal from 26 federated sites. Some items of discussion may include suggesting mechanisms for feedback for siteDB, designing automated updates to siteDB, and linking to the siteDB portal from each LTER website. In addition, our web site users have found it difficult to navigate the first generation of LTER sites' websites to downloadable "payoff" data and other current and exciting research information. We would like to explore siteDB as a portal to "payoff" data, as well as other information. As a working group, we would like to come up with the next steps for siteDB by suggesting what other categories of information, query tools, and data may be centralized in this directory.

Description of the siteDB Input Tables
Survey done online (http://www.lternet.edu/sites/)
By: Eda C. Melendez-Colom (LUQ)
For the LTER Information Managers (IMs) Web Development Group
July 27, 2005

siteDB's URL: http://www.lternet.edu/sites/)
List of sites: See copy of page:

Profiles from LNO site database
Andrews LTER (AND) [homepage]
Arctic LTER (ARC) [homepage]
Baltimore Ecosystem Study (BES) [homepage]
Bonanza Creek LTER (BNZ) [homepage]
Central Arizona - Phoenix (CAP) [homepage]
California Current Ecosystem (CCE) [homepage]
Cedar Creek LTER (CDR) [homepage]
Coweeta LTER (CWT) [homepage]
Florida Coastal Everglades (FCE) [homepage]
Georgia Coastal Ecosystems (GCE) [homepage]
Harvard Forest (HFR) [homepage]
Hubbard Brook LTER (HBR) [homepage]
Jornada Basin (JRN) [homepage]
Kellogg Biological Station (KBS) [homepage]
Konza LTER (KNZ) [homepage]
Luquillo LTER (LUQ) [homepage]
McMurdo Dry Valleys (MCM) [homepage]
Moorea Coral Reef (MCR) [homepage]
Niwot Ridge LTER (NWT) [homepage]
North Temperate Lakes (NTL) [homepage]
Palmer Station (PAL) [homepage]
Plum Island Ecosystem (PIE) [homepage]
Santa Barbara Coastal (SBC) [homepage]
Sevilleta LTER (SEV) [homepage]
Shortgrass Steppe (SGS) [homepage]
Virginia Coast Reserve (VCR) [homepage]
★LTER Network Office (LNO) [homepage]

- Contact info. for all sites
- Coordinates for all sites
- Site links
- Grant information for all sites
- LTER Sites Slide Show

siteDB Themes:

1. **Site**
2. **Climate**
3. **Vegetation**
4. **Regionalization**
5. **Soils**

siteDB Input Tables and their categories (forms) in user's interface:

1. **Site** (General Information)
    a. **Description**
    b. **URL**
2. **Sub-site** (Research Locations)
    a. **Locations**
    b. **Class**
    c. **Abiotic parameters**
3. **Research Theme** (Research Themes)
    a. **Theme description (?)**

Actual Menu of Input Forms:

General Information
Research Locations
Research Themes
Contacts
Affiliations

- **General Information Input Table Entries (a "*"denotes is a required entry):**

***Site Name:**
Synonyms:
Grant Number(s)
***Short Description: (A short 1 Paragraph description for display on the web.)**
Full Description:
***Short history:**
***Short Research Topics: (A short 1 Paragraph description of the research topics for display on the web.)**
Full Research Topics:
Site Administrative Contact Information

***Address 1:**
Address 2:
Address 3:
***City:**
***State: Select a State Alabama Alaska Arizona Arkansas California Colorado Connecticut Delaware District of Columbia**
Florida Georgia Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana

Maine Maryland Massachusetts Michigan Minnesota
Mississippi Missouri Montana Nebraska Nevada New Hampshire New Jersey New Mexico New York North Carolina North Dakota Ohio
Oklahoma Puerto Rico Oregon Pennsylvania Rhode Island South Carolina South Dakota Tennessee Texas Utah Vermont Virginia
Washington West Virginia Wisconsin Wyoming
***Zip:**
***Email:**
***Phone:**
Fax:

Site Website Information

***Home Page:**
Site Image URL:
Data URL:
Personnel URL:
Bibliography URL:
Education URL
Information Management URL:
RSS News Feed::
(An RSS News feed is a Format for syndicating News from you site. For more information on RSS please read the "What is RSS?"
article from the O'Reilly Network. Please contatc the LTER Network Office for information on setting up an RSS News feed for your site)
Newsletter:

Location of LTER Site

***Latitude: + - (Decimal Degrees min. 3 decimal places)**

***Longitude: - + (Decimal Degrees min. 3 decimal places)**
(These coordinates will refer to a point on a map that is representative of your LTER site. In some casese it can represent administrative
office at a University. To enter more detailed information on your site's actual research locations please go to the "Research Locations"
page and enter coordinates for your research locations there.)

- **Research locations Input Table (one for the primary site, and one for each sub-site) Entries:**

Asterisk (*) denotes a required field
***Research Site Name:**
Parent ResearchLocation:

Type: Primary Met Station Field Station Waters hed General Laboratory Flux Tower GFL Site
Site Classification(s):
***Description:**
Research site URL:
Research Site Image:
Director:
Steward:
Contact Name:
Contact email:
Area (HA):
Disturbance regime:
Location (fixed point)
Latitude: + - (Decimal Degrees min. 5 decimal Places)
Longitude: - +
Location (Bounding Box)
Latitude 1: + - (Decimal Degrees min. 5 decimal Places)
Longitude 1: - +
Latitude 2: + -
Longitude 2: - +
Elevation
Low elevation (m):
Mean elevation (m):
High elevation (m):

- **Research Themes Input Table (one for each Theme, i.e. study topic) Entries:**

Research Themes for <site>
Climate
Temp High: Temp Mean: Temp Low:
Windspeed Max: m/s Windspeed Mean: Windspeed Min: m/s
Preciptation Mean: dim_parm: dim_stat:
Snow Day: rel_parm: rel_stat:
Sunshine Day: aet: pet:
Water Balance:

Geology
Available Soon

- **Contact Input Table Entries:**

Update Contacts for:
Contact Type Contact Action
Lead Principal Investigator (LPI):
Primary Contact:

Education Contact:
Information Manager (IM):

Lastname: **Show all from my site**

- **Affiliations Input Table Entries:**

Update Affiliations for: <site>
Select an Organizational Affiliation AERC - Association of Ecosystem Research Centers ANL - Argonne National Laboratory ANL - Argonne National Laboratory ASU - Arizona State University BAS - British Antarctic Survey BYRD - Byrd Polar Research Center - The Ohio State University CAEC - Center for Analysis of Environmental Change CAES - Colorado Agricultural Experiment Station CDNP - Chihuahuan Desert Nature Park CERN - Chinese Ecosystem Research Network CQS - Center For Quantitative Sciences CRN - Canopy Research Network CSU - Colorado State University CU - Cornell University CU - University of Colorado CWM - College of William and Mary CWRS - Center for Water & Restoration Studies DOE - Department of Ecology DUKE - Duke University ECU - East Carolina University EMAP - Environmental Monitoring and Assessment Program (EPA) ENP - Everglades National Park EPA-EMAP - EPA-Environmental Monitoring and Assessment Program ERG - Ecosystem Research Group ESA - Ecological Society of America ESC - The Ecosystems Center ESRI - Environmental Systems Research Institute, Inc. FAB - Florida Audubon Society FGDC - Federal Geographic Data Committee FIU - Florida International University GATECH - Georgia Institute of Technology GCTE - Global Change Terrestrial Ecosystem Project GTOS - Global Terrestrial Observing System GTOSNPP - Global Terrestrial Observing System NPP Project HBOI - Harbor Branch Oceanographic Institute HGP - Human Genome Project HU - Harvard University IES - Institute of Ecosystem Studies (Millbrook, NY) IGBP - International Geosphere/Biosphere Program ILTER - International Long Term Ecological Research INSTAAR - Institute of Arctic and Alpine Research - University of Colorado ISEM - International Society for Ecological Modeling ITE - Institute of Terrestrial Ecology (UK) IU - Indiana University JGOFS - Joint Global Ocean Flux JWJERC - J.W. Jones Ecological Research Center KSU - Kansa State University LANL - Los Alamos National Laboratory LDEO-CU - Lamont-Doherty Earth Observatory-Columbia University LMER - Land-Margin Ecosystem Research Program LSU - University of Louisianna LTER - Long Term Ecological Research LTREB - Long-Term Research in Environmental Biology LTSS - Long Term Studies Section of the ESA MAB - Man and the Biosphere Program MAREO - Mid-Atlantic Region Ecological Observatory MBL - Marine Biological Laboratory (Woods Hole) MSU - Michigan State University NASA - National Aeronautics and Space Administration NBS - National Biological Survey NC - Nature Conservancy NCAR - National Center for Atmospheric Research NCEAS - Center for Ecological Analysis and Synthesis NCGIA - National Center for Geographic Information and Analysis@UCSB NCSA

- National Center for Supercomputing Applications
NMSU - New Mexico State University NONE - No Organizational Affiliation NPS - National Park Service NRCS - Natural Resources Conservation Service
NREL - Natural Resources Ecology Laboratory (CSU, Ft. Collins) NSF - National Science Foundation NSF-BIR - Biological Instrumentation and Resources
OBFS - Organization of Biological Field Stations ODU - Old Dominion University
OHSU - Ohio State University ONRC - Olympic Natural Resources Center
ORNL - Oak Ridge National Laboratory OSU - Oregon State University PDX - Portland State University PFEL - Pacific Fisheries Environmental Laboratory
PNW-AND - Pacific Northwest Forest Experimental Station - HJ Andrews PNWL - Pacific Northwest Laboratories PORG-M - Polar Oceans Research Group - Montana
PRBO - Point Reyes Bird Observatory RU - Rutgers University S2000 - Sequoia 2000 SAML - Southern Association of Marine Laboratories
SBI - Sustainable Biosphere Initiative SEEK - Science Environment for Ecological Knowledge SERC - Smithsonian Environmental Research Center
SFWMD - South Florida Water Management District SIO-UCSD - Scripps Institution of Oceanography - University of California at San Diego
SKIO - Skidaway Institute of Oceanography SML - Southern Marine Laboratories
SNEP - Sierra Nevada Ecosystem Project SREL - Savannah River Ecology Laboratory STRI - Smithsonian Tropical Research Institute SU - Syracuse University SWFSC - Southwest Fisheries Science Center TAM - Texas A&M University UAF - Univ of Alaska - Fairbanks UB - University of Bristol UBC - University of British Columbia UC - University of Colorado UCB - University of California Berkeley UCSB - University of California at Santa Barbara UGA - University of Georgia UH - University of Houston ULL - University of Louisiana at Lafayette UM - University of Minnesota UM - University of Miami UNC-W - University of North Carolina at Wilmington UNEP - United Nation Environmental Program UNH - University of New Hampshire UNM - University of New Mexico
UPR - UNIVERSITY OF PUERTO RICO UPRRP - University of Puerto Rico Recinto de Rio Piedras USC - University of South Carolina USDA - US Department of Agriculture USF - University of South Florida USFS - US Forest Service USFWS - US Fish and Wildlife Service USGS - US Geological Survey USGS-EROS - US Geological Survey - Earth Resources Observing System
USGS-GDCNM - USGS Geographic Data Coordination - National Mapping Div.
UUN - Utrecht University, The Netherlands UVA - University of Virginia
UW - University of Washington UW - University of Wisconsin VCU - Virginia Commonwealth University
VIMS-CWM - Virginia Institute of Marine Science - College of William and Mary

Add an organization to this list

Affiliation Action
US Forest Service Delete

Cross Table (See "Site Input table" sheet)

	Entry Item	Usage %
1	*Site Name:	100
2	Synonyms:	100
3	Grant Number(s)	96.2
4	*Short Description: (A short 1 Paragraph description for displ	100
5	Full Description:	88.5
6	*Short history:	100
7	*Short Research Topics: (A short 1 Paragraph description of	100
8	Full Research Topics:	88.5
9	Site Administrative Contact Information: *Address 1-3:	100
9b	Site Administrative Contact Information: *City:	100
9c	Site Administrative Contact Information: *State:	100
9d	Site Administrative Contact Information: *Zip:	100
9e	Site Administrative Contact Information: *Email:	100
9f	Site Administrative Contact Information: * phone	100
9g	Site Administrative Contact Information: fax	88.5
10	Site Website Information: *Home Page:	100
10b	Site Website Information: Site Image URL:	100
10c	Site Website Information: Data URL:	92.3
10d	Site Website Information: Personnel URL:	96.2
10e	Site Website Information: Bibliography URL:	92.3
10f	Site Website Information: Information Management URL	57.7
10h	Site Website Information: RSS News Feed	3.8
10i	Site Website Information: Newsletter	30.8
10j	Location of LTER Site: *Latitude: + - (Decimal Degrees mir	100
11	Location of LTER Site: *Longitude: - + (Decimal Degrees n	100
Usage AVG.		89.4

Site Input	1	2	3	4	5	6	7	8	9	9b	9c	9d	9e	9f	9g	10	10b	10c	10d	10e	10f	10h	10i	10j	11
AND	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*
ARC	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*			*	*
BES	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*
BNZ	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
CAP	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*
CCE	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*		*					*	*
CDR	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*	*
CWT	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
FCE	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
GCE	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*
HFR	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
HBR	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*				*	*
JRN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*
KBS	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*				*	*
KNZ	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*	*
LUQ	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*
MCM	*	*		*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*				*	*
MCR	*	*	*	*		*	*		*	*	*	*	*	*	*	*	*							*	*
NWT	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
NTL	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
PAL	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
PIE	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*				*	*
SBC	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*				*	*
SEV	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*		*		*	*
SGS	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			*	*
VCR	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*				*	*
LNO	*	*	*	*	*	*			*	*	*	*	*	*	*	*	*		*			*	*	*	*
PercentTot	100	100	96.3	100	88.9	100	96.3	85.2	100	100	100	100	100	100	88.9	100	100	88.9	96.3	88.9	55.6	7.4	33.3	100	100
% Exclude	100.0	100.0	96.2	100.0	88.5	100.0	100.0	88.5	100.0	100.0	100.0	100.0	100.0	100.0	88.5	100.0	100.0	92.3	96.2	92.3	57.7	3.8	30.8	100.0	100.0

Design of Query Interface to LTER Catalog: Controlled Vocabularies

One of the challenges facing LTER and external researchers in discovering data from LTER sites is inconsistent application of keywords. A researcher interested in carbon dioxide measurements must search on "Carbon Dioxide," "Carbon Dioxide Concentration", "Carbon Dioxide Emission", "Carbon Dioxide Evolution", "Carbon Dioxide Flux", "Carbon Dynamics", and "CO2" (all from the August 2005 Data Table of Contents). Moreover, additional key terms such as "gas flux" coupled with "carbon" may indicate data of interest. A study of keywords several years ago indicated that the modal number of LTER sites the typical keyword was used at was 2!

The problem is even more severe for less specific concepts. In principle, researchers interested in global change would be interested in carbon dioxide emission and uptake. Yet, a search on "global change" in DTOC gets only 5 "hits" with all but one from a single site and of these datasets, none included carbon dioxide measurements. A similar search using the KNB Metacat system found 17 datasets as compared to 33 hits for "carbon dioxide" (in principle one would expect the broader topic to get more, not fewer, hits).

The approach used by NASA's "Global Change Master Directory" has been to develop a hierarchical system of keywords and key terms which they characterize as "Category", "Topic", "Theme", "Variable" and "Keyword". The four highest levels of the hierarchy are based a controlled vocabulary of specific terms. The lowest level (keyword) allows researchers to pick their own terms. You can see their list at: http://gcmd.nasa.gov/Resources/valids/gcmd_parameters.html . This hierarchical system allows a "browsable" page where researchers burrow down from the most general terms to the most specific (http://globalchange.nasa.gov/).

The Oak Ridge National Laboratory Distributed Active Archive Center does not support keywords per se, but instead structures searches around project, investigators, sensors, parameters and sources (http://daac.ornl.gov/cgi-bin/search/asearch.pl) with a relatively limited number of key terms in each group (usually <200). Additionally, they support searches of free text or specific document parts including keywords, across multiple data centers (http://mercury.ornl.gov/ornldaac/).

We see several possible approaches to improving the search interface for LTER and hope the working group will be able to identify additional approaches.

1. List and reconcile existing keywords by creating a "look up table" of synonyms
2. Adopt or modify an existing list of keywords and recode LTER data sets using that list
3. Adopt or develop an set of hierarchical keywords or even an ontology (for a good description of ontology creation and use see: http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf).

We have done much to answer the challenge we set for ourselves way back in Spokane to implement a structured metadata standard. During the past few years we have learned a lot about the strengths and weakness of EML. This workshop gives us an opportunity to address a specific shortcoming with tools that will enhance our preparedness for adopting more advanced informatics solutions into the LTER. One clear shortcoming is the poor handling of data semantics. We have provided a great deal of structured elements to enter information on unit, domain, storage, precision etc, that do us very little good without the simple information on what IS being observed – that resides in a text description that references no common knowledge standards. Another flaw is that EML is a system designed to describe data from what one might think of as an archivist's, or manager's, point of view. It is structured hierarchically by datasets, entities, and attributes – the units by which we store and manipulate data. However, these are not always familiar to the user community we expect to use our data – a very plausible inqury might read: do you have any DOC data? What format is it in? How do I get it? The inquiry has progressed in just the opposite order of our management hierarchy , beginning with the measurements themselves, then to the file format and finally to the dataset information. Recent efforts to locate datasets suitable for a 25th anniversary volume on long term trends in LTER underscored this. Finally, EML does attempt to express some common semantics through controlled vocabularies for certain elements, but these have posed significant maintenance issues related to how they will be extended and updated.

In this workshop we will look at the potential for centrally shared ontologies as a means for providing the semantic definitions of data and related concepts that we currently lack and enable us to present search and access interfaces to our data that have a measurement-centric rather than dataset-centric view. Unlike many top-down approaches to building ontologies, we are focusing here on a bottom-up. We wish to create a system that minimally stores in a commonly accessible place, full semantic descriptions of each measurement that appears in one or more LTER datasets. Future research by SEEK and other projects will undoubtedly be able to lever this resource by adding in richer knowledge about these variables. But getting them in one place with one common identifier should be a feasible and productive first step. That said, our charge for the working group will be to evaluate:

1) Are there substantial suites of monitoring observations from LTER sites that could be described with a smaller number of shared class descriptions? Which are they? What miminally is needed to provide a useful definition?

2) Are there existing systems of measurement description that can be adopted or emulated? eg EPA Storet.

3) How do we adopt this technology to enhance what now do with EML? How do we decide what information is most usefully recorded in a dataset-bound EML document and what is more usefully stored in some central, reusable context? How do we reference shared ontology concepts from EML documents?

4) Given what we know about completing EML, what strategies should we pursue in introducing central attribute descriptions and how should they be populated and managed.

5) How do we expect this information to benefit us? What sort of features would we expect to be able to implement in our information systems?

While I think it will probably be instructive for us to actually create a few ontology descriptions for some very common core measurements, the product for

the working group will be to define a specification and challenge for developing an LTER knowledge base of measurement descriptions, much as the Spokane IM meeting produced the charge that launched EML.

Friday AM

1. Introductory remarks and questions
    1. mark Sh. OWL structure - class, subclass
    2. peter – topdown vs bottom up. Filling in the 'leafs' that map directly to whats in the dataset
        1. matt – bottom up needs some common framework – agrrement on what information is required for bottom level leaves.
        2. more frequent development and evaluation cycles
        3. mark Sv. - what are the fundamentl concepts of ontologies do we need to do this.
        4. karen – we need to think carefully about how this can be implemented given the experiences of GEON.
        5. peter – adding what goes into em-attribute into an ontology
            1. matt – we just need a standard class that says what a temperature measurement is then link our soil temp measurement to that concept.
2. Agenda items
    1. Morning – expectations & goals
        1. what do we want ontologies to do for us?
        2. are there common measurements within lter?
        3. what information is needed to describe an attribute?
    2. afternoon - Implementation issues
        1. common framework
        2. user interfaces both for query and submisssion
        3. EML integration
3. What do we want these to do
    1. linking data to fundamental concepts.
        1. do we map to a common set of mid-level concepts
        2. do we all enter each measurement as it is in the database.
        3. wade – can we identify some areas where we can work out more detail in some areas, but leave it open for others.
    2. bill – set priorites that will enable some rapid delivered products – sim to geon.
    3. dealing with mixed concepts in variable description – soiltemp vs temp+soil|air
        1. todd – registering each tag is useful for practical reasons
        2. margaret – using GCMD vocabularies in keywords
        3. wade – atttribute descriptor super type – conceptual parent like 'temperature' . dropped because it wasnt integrated into applications.
        4. ferdinando - having some system to manage the redundancy – search repository to find what your looking for and then if not found, then add your own.
        5. lynn- instiuted common attribute names. 'intent' – nutrient >nitrogen>concentration of nitrate.
        6. linda – registering a dataset - “eh” is same as some other measurement. how do we standardize?
        7. ferdinando – having a set of established leaf concepts will guide future data definition and registration.
        8. judy – we have attributes and we have terms. the onology allows defining relationhips among themselves
        9. whats the strategic plan?
            1. inventory core measurements at sites? what measurments are sites responding to Trends?
                1. evaluate for redundancy
                2. evaluate for semantic 'complexity'

        2. simple, clear, defined, concise
    10. reusing descriptive information
        1. templates?
    11. enabiling discovery
    12. support integration
    13. what goes in attibute definitions
        1. domain? matt says no because domans can be meaningfully constrained by context ie, body length of rodents.
        2. we may have to not worry about
4. Implemtations concerns
    1. referencing common vocabularies imples some form of stustainable, versioned systems.
        1. classes have to exist forever
        2. versions
    2. resource needs – increases needs to maintain
    3. dealing with storage and versioning in partnership with seek?
    4. emery – we need to try it and see what we do.
5. Pilot pilot studies
    1. water chemistry
        1. found commn variables eg DOC.
        2.
    2. biodiversity
        1. konza – five biomass columns – subclasses of live biomass. live biomas woody
        2. sampling was an important but difficult concept to evaluate.

sweet.nasa.gov

LTER Annual IM Meeting, Montreal, Canada, Friday Aug 5, 2005
Working Group Session II, Group 4: Web Site Design Recommendations

This working group was formed to produce recommendations for sites developing next generation websites and tools. These recommendations also will be referenced in the IM review criteria document. Web users and developers have identified challenges in navigating complex web sites, accessing usable data and metadata, reporting current research information and exciting news, and portraying that they are part of a larger LTER network. This group has already surveyed the IM committee to describe our web site users and what role they have played in contributing to the design of web sites within the community (http://intranet.lternet.edu/committees/information_management/im/webdev_questions_summary.pdf). Our recommendations will include, suggested labels or names for categories of information that would be expected to contain similar content from site to site and design elements and links that communicate a site is part of a greater network.

Dictionary Process: Unit Registry Prototype
Working Group 5 Description

Well specified units are at the heart of measurement comparability. Today's data integration efforts highlight a range of unaddressed and unresolved unit definition issues involving syntactic and semantic ambiguities and conflicts. A Unit Registry Prototype that addresses some of these issues will be demonstrated (see http://fire.lternet.edu/customUnit/). It enables sharing unit names, types, and definitions while introducing the concept of multiple levels of scope, ie site, working group, community, and domain. The working group discussion will range from sharing how units are now being handled at the site to how dictionaries might facilitate local as well as cross-site needs.

LTER Information Manager Meeting 2005

Working Group 5 Report
Baker, Servilla, Yarmey, Powell, O'Brien, Sheldon, Haber, Millerand, San Gil
Unit Dictionary Working Group
August 5, 2005 (1:15-3:15) Montreal

The Unit Dictionary Working Group was formed in order to consider and broaden the work of a Unit Registry Design Team that met periodically over many months prior to the LTER IM Meeting. The Design Team started with an understanding that the unit dictionary, distributed with EML and based on SI with base and derived units in STMML format, is useful and would be of greater benefit to users if developed further and updated. Out of this common view came the notion of a unit registry that took into account the added value of having a mechanism for exchanging and comparing units across sites. Recognized needs included:

(1) developing a unit registry, in effect migrating the unit dictionary outside EML
(2) building a web application unit registry viewer
(3) sharing the unit registry schema and code
(4) initiating a unit registry collection in order to gain familiarity with extent of project
(5) opening up the 'scope' term from EML or custom to include other designations such as site_name, WG, and LTER, an indicator of scope of acceptance starting at the site level and migrating to the domain (EML) level.

The prototype developed is currently available online http://fire.lternet.edu/customUnit. The web site highlights 'Motivation and Process' as well as 'Products":

> *Well specified units are at the heart of measurement comparability. Today's data integration efforts are highlighting a range of unaddressed and unresolved unit definition issues involving syntactic and semantic ambiguities and conflicts. The community repository or "unit registry" presented here addresses some of these issues by sharing unit names, types, definitions, and forms while introducing the concept of site-working group-community-domain scope. The initial goals for our design team were three-fold: 1) preparation of a* ***prototype*** *application to enhance discussion at a Dictionary Working Group at the August 2005 LTER Information Manager meeting, 2) establishment of a* ***community process*** *that engages and benefits participants, and 3) development of a* ***dictionary process*** *that captures the migration of units through multiple levels. This prototype represents an LTER site-network collaborative design effort to meet community needs by creating a mechanism for locating units compliant with the EML standard, for bringing together local solutions, and for prompting cross-site discussion of units.*

These goals were met by the time of the IM Meeting where the prototype was demonstrated the first evening to all participants and again in our working group. Interested sites were invited to join the next unit registry design effort. The code is available from the online CVS repository http://cvs.lternet.edu.

Design Team communications included conference calls held weekly or biweekly over many months. Participants included Karen Baker (PAL/CCE), Mark Servilla (LNO), Wade Sheldon (GCE), Linda Powell (FCE), Margaret O'Brien (SBC), Inigo San Gil (LNO), Lynn Yarmey (CCE), Shaun Haber (PAL), and Florence Millerand (PAL). A working web site served as a communication mechanism to augment use of email and to provide an ordered archive of materials. The group worked collaboratively on a unit repository schema which was updated and circulated periodically as well as posted on the unit repository website.

The Working Group began with a round-robin focused on sharing experiences, understandings, questions, and local practices. A short background was provided on lexicography, dictionaries and change by Lynn Yarmey, on non-physical units by Margaret O'Brien and Linda Powell, on technical factors by Mark Servilla, and on community factors by Inigo San Gil. In addition, Wade Shelton summarized with an ontology-readiness eye to forward planning.

The working group identified some next steps:
1) consider new units and new unit types
2) allow duplicated units as customUnits in EML; consider duplication as a mechanism to help resolve differences
3) create a web input submission form for unit submission
4) with unitType a tie to attributes, consider range of attribute dictionary approaches
5) consider how to frame the units issue to be of interest to an environmental scientist
6) call for including new participants with an interest in this topic

LTER Knowledgebase for Attribute Descriptions

Peter McCartney (CAP)

CAP LTER
Central Arizona – Phoenix Long-Term Ecological Research

Goals

- Develop Strategic plan for implementing central knowledge base of attribute description ontologies in LTER
  - Mission
  - Goals
  - Objectives

- Are there substantial suites of monitoring observations from LTER sites that could be described with a smaller number of shared class descriptions? Which are they? What miminally is needed to provide a useful definition?

- Are there existing systems of measurement description that can be adopted or emulated? eg EPA Storet.

- How do we adopt this technology to enhance what now do with EML? How do we decide what information is most usefully recorded in a dataset-bound EML document and what is more usefully stored in some central, reusable context? How do we reference shared ontology concepts from EML documents?

- Given what we know about completing EML, what strategies should we pursue in introducing central attribute descriptions and how should they be populated and managed.

- How do we expect this information to benefit us? What sort of features would we expect to be able to implement in our information systems?

Rationale

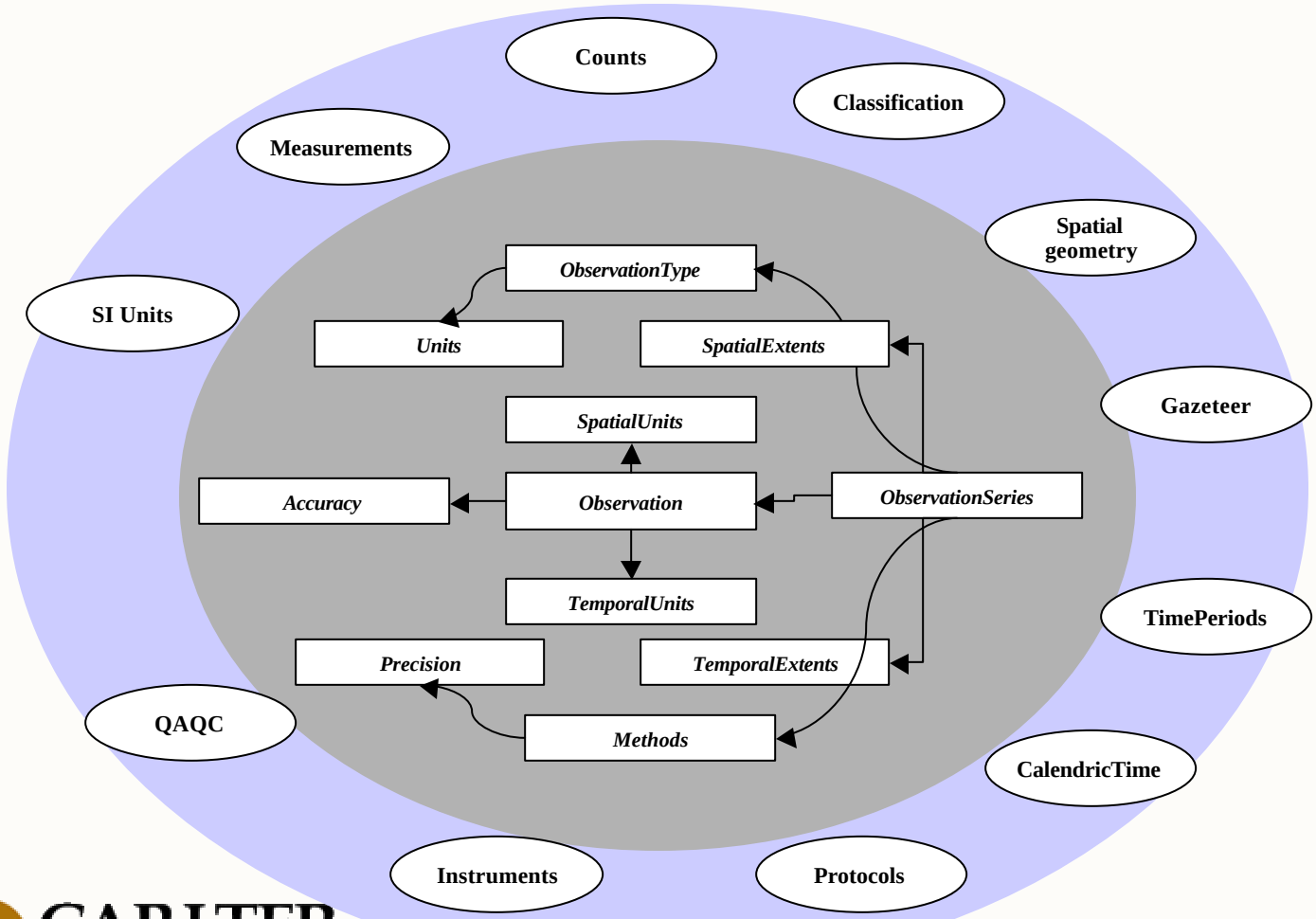
- Enhancing EML metadata
- Improved data access interfaces
- Supporting data integration research

EML Issues

- Poor control over enumerated vocabularies – units, spatial rep, file formats, etc
- attribute descriptions rich in ancillary info yet lack core semantics
- Shared attribute ontologies would provide comment reference for observation descriptions and resolve many extensibility isses with EML

User Interface issues

- Trends project
  - annual trends in core measurements from each LTER site
  - Identified conflicts between how scientists approach data search and how archives are organized – lumping and splitting, top-down, core/non-core
- Discussions between Brunt, McCartney, Peters on restructuring data archives from an observation oriented perspective
- Shared Attributes would enable discovery and query interfaces that emphasize observations



CAP LTER
Central Arizona – Phoenix Long-Term Ecological Research

Integration Research

- SEEK
  - ontologies to link data concepts with analytic and processing concepts
  - search and discovery
  - Integration
  - Shared attribute for LTER data provide a basis for operationalizing these tools

Community Process and Standards Implementation

The LTER Information Management community has been engaged in generating and implementing information management technology appropriate for network level data discovery, access, and integration for some time now. Part of this development has been conducted in partnership with other groups. This working group will provide an opportunity for information managers and developers to dialogue and reflect together on past development processes. We will use the community processes in EML design/development/deployment/enactment as a focal point. The discussion can inform upcoming EML revisions and future network projects. Products of the working group will include the accumulation of experiences of the participants with standards, distillation from these experiences of some principles and critical questions to guide the LTER IM community and its partners in future projects, and a conceptual diagram that identifies important steps and processes within the cycle.

Joint session

1. discussion points
    1. Whats the mission
        1. what is the overlap between the three working groups
        2. are there clear distinctions?
        3. are these three discreet products?
2. Whats the process
    1. what is the impact with EML and how do we coordinate these groups with EML development?
    2. articulation with research projects (SEEK, SWEET, GEON, etc)
    3. do we create a task force – or three?
    4. some sort of best practices type form of communication process.
    5. whats the time frame
3. how do we actually build these tools
    1. technical issues – formats, software for carrying out at least a pilot project
    2. do these require distinct technologys or can they share a common framework/approach?
    3. practical goals – separate short term and long term goals.

Discussion

Short term goals

1. units
    1. shorterm goal – within a year?
        1. make bettter eml
        2. easier to define and reuse unit definitions with browsable lists
        3. mechanism for sharing and vetting?
    2. implementation
        1. integration between eml standard and metadata creation tools.
        2. how can we make this accessible to diversity of creation tools?
        3. modify EML schema to accommodate external reference to base units
2. vocabularies
    1. goals
        1. discover existing terms in use keywords, titles, etc.
        2. browsable hierarchy of search terms for a search interface
        3. query existing web service thessauri for additional term expansion in addition to maintaining ecology – unique terms
        4. enrich eml documents with a richer set of keywords.
3. attributes
    1. mine existing attribute descriptions from EML and catalogs for core monitoring data
        1. look for redundancy
        2. focus on observations not contextual or methodlogical qualifiers
        3. evaluate instances of 'contexualized' measurement names
    2. enable reusable content
    3. modify eml to accommodate referenceexternal ontologies

case study?

1. data integration?
        1. this might be better as a long term goal – involves other issues like methods, context etc.
2. discovery is a shorter term product goal
3. progression from dictionary, to thesaurus, to ontology
        1. simple un ordered to highly structured

        2. distictions between controlling vocabularies and expressing knowledge unbounded
    4. is everything we are doing ultimatly leading to an ontology type structured
        1. challenge is to determine what is in the shared pool vs what is defined in individual metadata/research contexts
    5.
    6. gazetteers as controled vocabularies

Implmentation issues

1. mining existing systems like gazetteers and online thesauri
2. creating systems that we ensure are mainainted and available

whats the relationship between these groups

1. linking between thesauri and attribute names
    1. duane – we want search terms organized in a hierachy using an ontology
    2. there may be a problem here related to the expansion to a large number of terms that might require some way to restrict what relationships are pursued
2. linking between attribute names and restricted sets of units
    1. how do we coordinate these two –
        1. units are properties of attributes
        2. STMML units have been imported into OWL

Process

1. work in different groups because we are at different stages of progress
    1. possibly share a common subdomain focus
2. each group needs to articulate with EML
    1. entering bugs in bugzilla
    2. monitoring/posting to EML-dev
3. relationship with seek
    1. unfilled position available to help with ontology development

2006 LTER All Scientist Meeting Planning (Ramsey, Kaplan)

The goal of this working group is to determine the number and style of information management workshops at the 2006 LTER All-Scientist Meeting (ASM). A group of interested Information Managers will be formed to further the discussion started in this workshop. The working group will identify the general themes or goals for the information management workshops. From comments received from the 2003 ASM, information managers and researchers would prefer a more engaging style of workshops instead of presentations followed by questions and answers or discussion. The working group would also discuss the need (or not) of an Information Manager being on the LTER All-Scientist Meeting Organizing Committee. Please refer to the 2003 LTER All-Scientist Meeting web site for background information from the last ASM.