DATA INFORMATION AND MANAGEMENT

INFORMATION MANAGEMENT

The goal of the PIE LTER data and information system is to provide a centralized network of information and data related to the Plum Island Sound Estuarine Ecosystem and its watersheds. This centralized network provides researchers associated with PIE-LTER access to common information and data in addition to centralized long-term storage. Data and information are easily accessible to PIE-LTER scientists, local, regional, state partners and the broader scientific community. Researchers associated with PIE-LTER are committed to the integrity of the information and databases resulting from the research.

PIE-LTER information and databases are stored on a Microsoft Windows server with level 3 raid, which is backed up on tape nightly. Once a month a tape is removed and stored in a separate building.

Public access to PIE-LTER data and information for the scientific community at large is provided through the PIE-LTER World Wide Web home page on the Internet at the following URL: http://ecosystems.mbl.edu/PIE. Near real time weather data are also available on our field station website, www.pielter.org. The PIE-LTER home page has been active since late 1998 and contains information on personnel, data, published and unpublished papers, reports and School Yard education. The data section is broken down into four sections consisting of Program Areas, Education and Outreach, Physical Characteristics and Database Links. PIE maintains an internal database archive of datasets from which the home page is updated annually. Datasets on our web site are updated more frequently as investigators add data. The organization of the PIE home page basically mirrors the internal database archive in nomenclature, which allows for easy updating of datasets.

DATA MANAGEMENT AND COORDINATION OF RESEARCH PROGRAMMATIC AREAS

The information management team consists of: Chuck Hopkinson (Lead PI), Joe Vallino (PI), Robert (Hap) Garritt (IM), Gil Pontius (PI) and additional research assistants associated with program areas. The team has the necessary leadership, knowledge and technical expertise for creating and maintaining the PIE LTER research information. Hap Garritt, a senior research assistant with The Ecosystems Center, MBL, has been the information manager (IM) since 1998 and has the responsibility for overseeing the overall integrity of the data and information system for PIE-LTER. Hap has 25 years experience in ecological research, an MS in Ecosystems Ecology and is very active in PIE LTER research. Hap's regular research activities involve him with the design and execution of many of the research projects, which allows for a smooth incorporation of data and information into the PIE database.

Individual researchers are responsible for providing data in each of the six core programmatic areas outlined in the PIE-LTER (Watersheds, Marshes, Planktonic Food Web, Benthos, Higher Trophic Levels and Synthesis). Several meetings each year provide each researcher the opportunity to communicate with the PIE information management team regarding the design of the specific research project and subsequent incorporation of data and information into the PIE-LTER database.

CONTRIBUTIONS OF DATA TO DATABASE

Researchers on the PIE-LTER are expected to follow the LTER Network data release policy defined on the LTER web page, http://lternet.edu/data/netpolicy.html. Research conducted using

the facilities of the PIE-LTER is expected to comply with the following policy: All researchers will provide digital copies of data to the data manager. Data files will include accompanying documentation files that will completely describe the data. We have migrated from a Microsoft Word metadata template to a Microsoft Excel spreadsheet template. The Excel template was developed by Jim Laundre, ARC LTER and has been adapted for PIE to allow for consistent metadata entry and subsequent conversion via a visual basic macro to XML structured Ecological Metadata Language (EML) according to EML Best Practices for LTER Sites. Individual researchers are responsible for quality assurance, quality control, data entry, validation and analysis for their respective projects. Researchers are reminded about contributions to the database several times during the year via email or during field sampling trips, in addition to announcements during our Annual Spring PIE-LTER All Scientists Meeting.

DATA ACCESSIBILITY AND TIMELINES

Researchers on the PIE-LTER have been and will continue to be encouraged to both publish and contribute data to the PIE-LTER database. It is recognized that investigators on PIE-LTER have first opportunity for use of data in publications but there is also the realization for timely submittal of data sets for incorporation into the PIE-LTER database. Data is typically posted on the WWW within one to two years and selected data is made available in near real time to promote ecological awareness of the local environment. PIE follows the data release policy for the LTER network that states:

"There are two types of data: Type I (data that is freely available within 2-3 years) with minimum restrictions and, Type II (Exceptional data sets that are available only with written permission from the PI/investigator(s))."

PIE data sets and information are easily accessible to PIE-LTER scientists, local, regional, state partners and the broader scientific community, as we have no registration requirements for either observing or downloading data from our WWW page, which results in unobstructed access to all PIE LTER databases. Access to PIE data on the WWW is accompanied by a metadata document, which requests (based on an honor system) those users of the data to notify the corresponding Principal Investigator about reasons for acquiring the data and resulting publication intentions. However it is possible for users to download data without sending notification. We believe that unobstructed access to our data will encourage users to browse our WWW page and become involved with our research.

On-line PIE LTER data set usage is represented in Table 7-1. Researchers at PIE do get requests via email and phone for particular data sets, but since we have no formal registration, we have no formal documentation of the requests. However, analyses of PIE Web server log files, after removing web spiders, crawlers and web hits not associated with browsing, indicates that our on-line data sets are viewed extensively. On average during 2006 PIE had more than 15,000 hits/month on our home page, more than 5,000 hits/month on our Programmatic Area data pages

NETWORK PARTICIPATION

The PIE LTER program participates in the annual LTER Information Managers meetings, contributes to network level databases of ClimDB, HydroDB, Personnel, Bibliography, Site DB, Metacat Data Catalog and Trends and has been involved with LTER Network EML workshops. Hap Garritt is on the LTER IM Executive Committee (2007-2010) and is a member of the Unit

Dictionary and GIS Information Manager working groups. Chuck Hopkinson is on the LTER Network Information System Advisory Committee (2005-2009).

CURRENT IM PROJECTS

We are currently updating our existing on line EML level 2.5 metadata to EML level 4-5 using the MS Excel based template. We are in the midst of adding extensive datasets from the Tidal Creek fertilization experiment TIDES project and from stations pertinent to PIE LTER watersheds available from the NOAA National Climate Data Center Weather Cooperative and the National Atmospheric Deposition Program. Development of a GIS information system for sharing PIE LTER GIS data has been an on going project for many years as we are attempting to bridge three GIS softwares (ArcGIS, IDRISI, RiverGIS). The current LTER IM GIS working group is also discussing a centralized shared platform for GIS information as many LTER sites need a better way of viewing available GIS information at the site level and network level. We are also in the midst of redesigning our web site, the third time in 9 years.

FUTURE OBJECTIVES

Large streaming datasets associated with short sampling interval (15 min) weather and water quality station data loggers will require us to develop a database system capable of managing multiple year aggregations of data. The LTER Network as a whole and other planned observatory networks (AEON, NEON, ORION) are also in the midst of brainstorming how to cope with the vast amounts of data that will be forthcoming with these new environmental observatory initiatives. PIE has been and plans to continue to be involved in environmental observatory associated workshops.

Table 7-1. Monthly use of the PIE LTER Web Site during 2006 summarized from analyses of the PIE web server log files. Use is represented as the number of times a web page has been accessed [a] (hits).

Year 2006	Hits on PIE Homepage		Hits on PIE Programmatic area data	
Month	All Hits	Non-MBL Domain	All Hits	Non-MBL Domain
JAN	15,427	14,346	5,365	5,066
FEB	13,692	12,946	4,652	4,597
MAR	16,681	15,532	6,003	5,703
APR	17,554	16,785	5,961	5,831
MAY	16,596	15,434	5,868	5,585
JUN	14,640	13,841	5,045	4,794
JUL	16,754	15,067	6,106	5,732
AUG	16,116	15,248	4,675	4,582
SEP	18,917	18,038	5,779	5,628
OCT	17,773	16,995	5,023	4,899
NOV	19,932	19,456	7,007	6,981
DEC	12,070	11,500	4,446	4,318
Annual Total	196,152	185,188	65,930	63,896

[a]Non-browsing activity web hits have been excluded using filters for spiders, crawlers and domains not representing normal browsing activity.