Santa Barbara Coastal Long Term Ecological Research (SBC LTER)

Information Management Overview and Plan

Margaret O'Brien
January 2015

EXECUTIVE SUMMARY

The primary responsibility of the Santa Barbara Coastal LTER Information Management System (IMS) is to handle data and information produced by our NSF grant, and from some collaborators. A secondary goal is to contribute to informatics projects in the LTER Network. The primary goal of our documentation is to inform interested audiences and to assure continuity. This document (the "IM Plan") serves as the overview and timeline, and its intended audience is a reviewer, either external or the SBC IMS advisory committee. It contains brief descriptions of all IMS components. Other types documentation are component-specific descriptions and schematics, and guides for IMS staff, assistants and scientists. Some documentation is available only to SBC members. SBC's IMS is closely integrated with the UCSB Marine Science Institute (MSI), and the information manager (author) also collaborates closely with IMS personnel from Moorea Coral Reef LTER (MCR).

Major changes to this document

This document was initiated in the current format in 2009, and the text is updated annually. Appendices may be updated more frequently. A summary of major changes in 2015 below.

Section 1. Inserted "Requirements of SBC IMS" as Section 1.2

Section 2.2 (Integration of data processing and data publication) and 3.2 (Metadata) - incorporates use SBC Metabase, formalized in 2014.

Section 3.10, Data Catalog - anticipated updates to the local catalog (in 2015), necessary because of changes in data package management at the network/community level, e.g., the PASTA system and DataONE.

Section 3.11, EML Data Query application – Text moved to Section 4.4, under "SBC Informatics Projects".

`

1. INTRODUCTION

1.1. Mission Statement

The primary objective of the Santa Barbara Coastal LTER Information Management System (IMS) is to facilitate diverse research and outreach goals by focusing on ease of access, data organization and integrity, and long-term preservation. Its general, our IMS is:

1. Cross-platform and largely based on Internet standards rather than on vendor-specific technology,
2. Integrated with laboratory science to keep participants up to date on changes and advancements, and
3. Modular, allowing us to incorporate the skills of a broad community, enhance components individually, and to integrate SBC activities with those of the LTER network and our other collaborators.

The SBC IMS currently meet or exceed all standards listed in the LTER Network's Review Criteria for LTER Information Management. We anticipate that IM standards of the LTER Network will continue to evolve with emerging technologies and information needs, and will maintain our leadership role in this area to ensure that the SBC IMS is well positioned to meet the future expectations for LTER IM.

1.2. IMS Requirements

The SBC IMS is expected to oversee the housing and security all SBC data, allow sharing among SBC participants, and to design, produce and catalog data products for broader dissemination. Additionally, the IMS handles contributions from the SBC to the LTER Network catalogs and some material for reports to the funding agency (NSF). Nearly all SBC information is accessible via the Internet, and so design and maintenance of website(s) also falls under the purview of the IMS. The SBC IMS is not responsible for hardware or system administration, as this is provided by MSI, and includes infrastructure for file, application, database and web servers, with appropriate support software, described in this document where appropriate.

1.3. IMS Documentation

SBC's IMS is documented at several levels. Some documentation is publicly available on the website (http://sbc.lternet.edu).

(1) A general Information Management (IM) Plan (this document). Its intended audience is a new employee, a reviewer, either external or the SBC IMS advisory committee. An IM Plan is required for all LTER sites. This document is occasionally available publicly, or by request. It contains appendices that may be updated more frequently than the IM Plan itself:

I. Description of the IMS hardware and software stack.

II. An inventory of data packages organized by management needs and classified according to both local and Network categories.

III. Timeline for improvements to the IMS components

IV. List of research collaborators and their relationship with the SBC LTER IMS.

(2) An IM Guide (Wiki), whose intended audience is the IM staff and assistants, and which assures continuity. The IM Guide is updated frequently, as needed, and requires a login.

(3) Schematics, descriptions, revision notes, etc., of individual system components as reference material for current and future IM staff. Component documentation is stored in dedicated directories in each project's repository, and so included with 'check-out', or can be browsed through the web after login.

1.4. Personnel

SBC's IMS is closely integrated with the UCSB Marine Science Institute (MSI.ucsb.edu) and the Moorea Coral Reef LTER (MCR.lternet.edu). SBC has a dedicated information manager located at MSI (Margaret O'Brien) with contributions from the project coordinator (Jenny Dugan), MCR information manager (M. Gastil-Buhl), and MSI IT personnel (Jim Woods, Brian Emery). We also collaborate with several other LTER sites on ad hoc projects, and the Ecoinformatics program at the National Center for Ecological Analysis and Synthesis (NCEAS.ucsb.edu/ecoinfo), also located at UCSB. Major data contributors designate research staff members to interact with the SBC information manager, and about 80% of researchers' laboratories are located at UCSB, chiefly MSI, the Earth Research Institute (ERI.ucsb.edu), and Bren School for Environmental Science and Management (Bren.ucsb.edu). SBC also employs occasional assistants or undergraduate students for directed tasks as funding permits. A history of SBC IMS contributors is kept in the IM Guide.

Communication between SBC scientists and Information Management is fostered by the SBC IMS Advisory Committee (IMSAC), which includes SBC co-PI John Melack (chair), the information manager and two rotating Investigators from different research fields (currently D. Reed and D. Siegel). The IMSAC establishes priorities for SBC's IMS activities and increases the scope of researcher involvement in the SBC IMS.

1.5. Policies

For data sharing and publication, SBC's policy is aligned with the LTER General Use Agreement and the Network's "Type I-II" designations, and the policy is included with data packages. Type I data are generally posted publicly within 1 or 2 years of collection, although some ongoing electronic data are available sooner and data requiring complex chemical analyses or data processing procedures may be delayed. For "Type II" data, our policy is that data will be described in the public catalog, but the tables require authentication before delivery. A policy of requiring authentication may reduce availability of basic package information in federated catalogs (such as DataONE), and so currently, no SBC packages fall into this category. In addition, SBC employs a "Type 0 (zero)" designation for data we have acquired from outside parties and for which Network policies do not apply (e.g. USGS stream flow). These data may be republished per guidelines from the original producer. SBC also has posted a website Privacy Policy which is aligned with those of the University of California (SBC's local internet provider and host institution) and the University of New Mexico and LTER Network (owners of the DNS registration). All policies are available on the website.

The SBC IMS's primary responsibility is to handle data produced by our NSF grants. As with all LTER projects, SBC leverages and/or collaborates with other research conducted in the Santa Barbara area, and in some cases, these associated projects also leverage the SBC data systems (primarily the file server). The LTER Network is considering a policy under which collaborative projects that make use of an LTER site's IM resources agree to publish their data with LTER data. In 2011, NSF proposals were required to include a data management plan, and SBC began assisting collaborators with this process. If these collaborators decide to make use of the SBC IMS, their data would be covered by this policy. A list of collaborative projects and their relationship with the SBC IMS is in Appendix IV.

1.6. IT Systems

SBC's holdings are stored in a networked directory system, and a user account is all that is required to view any data file. Write-access is limited to those responsible for data collection and maintenance. The

directories for incoming data are maintained separately from those for "final" data products that are intended to be shared between disciplines or to be published. With this system, data are available to all SBC members immediately, as well as for processing or publication. Our common data areas have been stable for several years so returning users will remain familiar with the structure, and the directory structure is published internally. SBC account holders are encouraged to use their home directories for work-in-progress to take advantage of regular backups. Home directories are not shared.

Our IT components include a software repository, issue tracking, file system backup and LDAP, all supplied by the Marine Science Institute (MSI). See Appendix I for information about backups and the software stack.

Table 0. Major SBC LTER information management systems features

Type	Feature	Specification
Website, catalogs and/or directories	[1]http://sbc.lternet.edu [1]Bibliography (searchable) [1]Personnel directory (searchable) [1]Data (searchable by creator, LTER core area, site keywords) [1]Signature datasets Research projects Sampling sites	 [2]EML [3]LDAP [2]EML EML [2]LTER-project XML [2]SBC-places XML
Datasets	177 datasets (290 data tables) available SBC's and the LTER Network Catalogs	[2]EML
Database	[2]SBC Metabase, Metadata exchange	[2]GCE Metabase2, PostgreSQL [2]EML
Servers and user accounts	Web, database and file systems with backup 159 user accounts.	[3]LINUX, [3]SAN, [3]ext4, rsync [3]LDAP
Code Repository, Issue tracking, Documentation	Code storage with revision control for data models, website components; issue tracking and help-queue; system documentation	[3]SVN, [3]OTRS, SchemaSpy, DokuWiki

Notes: [1]feature required by all LTER sites, [2]further described in text, [3]provided by UCSB Marine Science Institute

1.7. Definitions

Data package: data entity (or entities) and metadata. Data entities are most often tables.
Data package update: The addition of new data (and metadata) to a data package, usually an ongoing time-series.
Data package maintenance: Enhancement of metadata or data to improve presentation or usability, or to keep the package current with standards or best practices.

2. DATA

2.1 Data types

Data come from diverse scientific endeavors in a variety of habitats including terrestrial/riparian, streams, beach, reef and ocean. Examples of measurements are in Table 1, and a complete inventory of

current public and anticipated data products is in Appendix II. As of this writing, 156 packages are publicly available and described in Ecological Metadata Language (EML), the metadata exchange format used by the LTER Network. Public data holdings comprise 291 data entities (e.g., data tables, KML files, images). Approximately half the inventory is ongoing time-series, in which the package is regularly reviewed and data added (i.e., updated, see definitions). All data packages are maintained, i.e., kept up to date with current standards and practices.

Table 1 Data types and measurements managed by SBC's information management system

Discipline	Representative measurements
Hydrology and meteorology	Stream discharge, precipitation
Oceanography	Moored and profiled hydrography (CTD), currents (ADCP), optics, swell
Biogeochemistry	Major nutrients, cations, particulate carbon and nitrogen, and pigments
Populations and community structure	Algae and animal survey data in fixed transects and experimental plots
Ecosystem processes	Rates of elemental flux, primary production (various methods), stable isotopes
Genomics	Organizational Taxonomic Units (OTU), microsatellite markers
Remote sensing	Kelp canopy biomass from Landsat, AVIRIS (anticipated)

2.2. Integration of data processing with data publication

SBC has deliberately not chosen a system where datasets are "submitted" by scientists and published by the IM staff. Instead, data are co-managed by the information manager and the data owners (i.e., investigators and their research staff), and wherever possible we integrate data publication with data processing. Integration is essential for datasets that are designated as "ongoing" since scientific personnel are the source for knowledge about changes to sampling protocols and data output. This management style is complicated, and requires coordination among diverse scientific domains, measurement types and laboratories. The IMS accommodates several data processing patterns and software choices (e.g. SAS, Matlab, and MS-Excel). Scientific personnel are trained as necessary in informatics concepts (e.g., data table design, SI units), programming practices, and use of the shared file

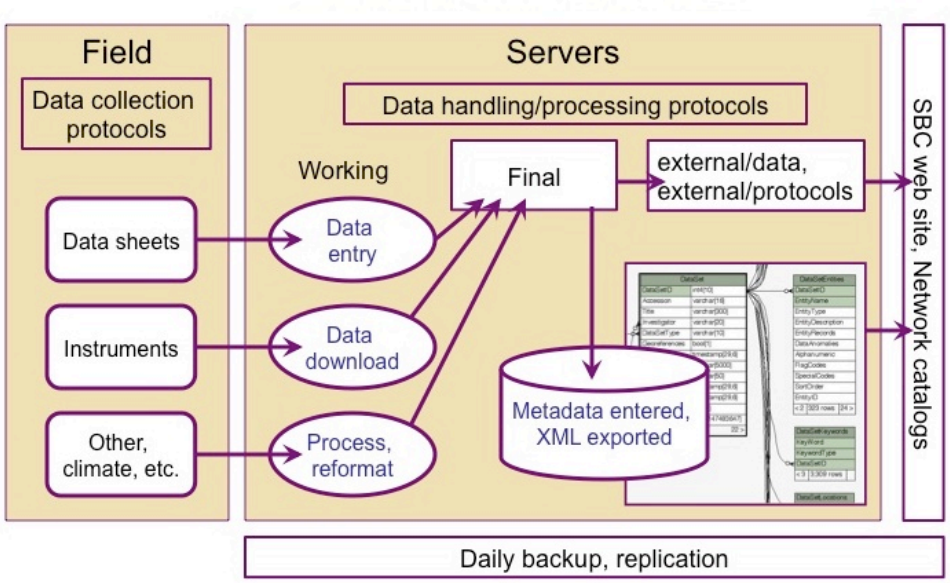

Figure 1. SBC LTER IMS. Integration between data related activities and file system and eventual data distribution in Network and catalogs.

server and the "Final" area. Structured metadata in the RDBMS and XML export is the handled by the information manager and trained assistants. Coordination and training is the responsibility of the information manager. Figure 1 shows the general pattern of data collection to data package distribution, and the IMS is involved as soon as data is uploaded to the server.

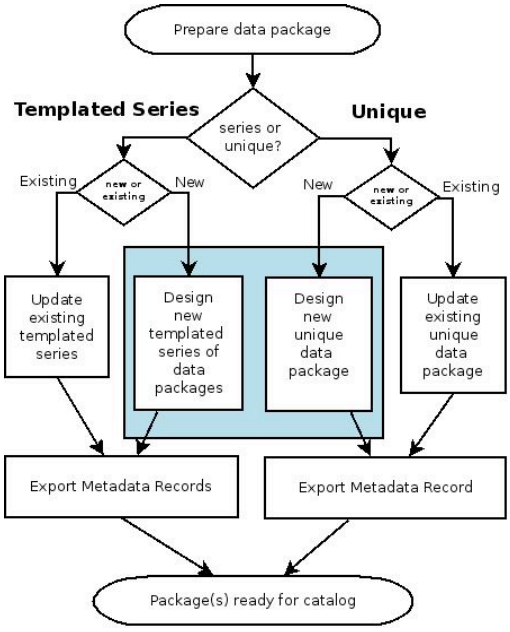

Figure 2. Pathways of data package preparation: "template" and "unique". The colored box indicates design process necessary for all new packages; details are in other documentation.

There are currently two major pathways for data publication: (1) "template", in which data are published as the last step of data processing using templates and (2) "unique", where scientific staff members produce the data entities, and metadata records are built by the information manager after review and metadata entry. Figure 2 shows the pathways for data package preparation.

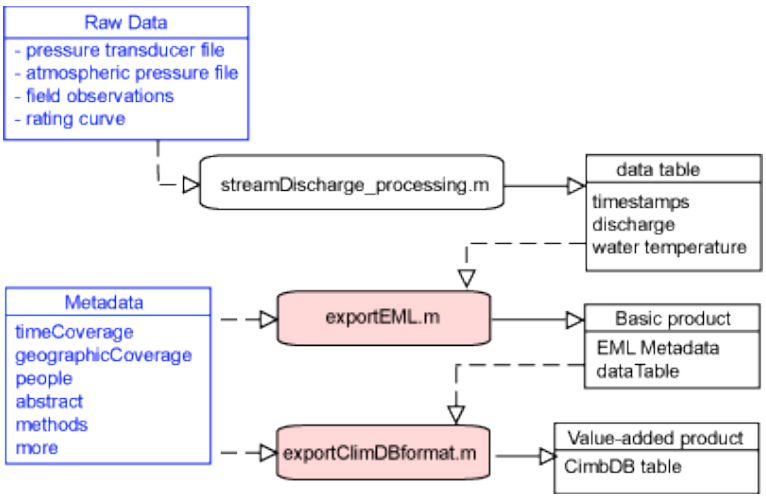

Figure 3. Data procerssing flow for SBC's stream discharge data. Components contributed by the IM team are in pink boxes. Data inputs are in blue, and data products are on the right.

The "template" pathway, in which data are published as the last step of processing is appropriate for data in which the format rarely changes and the processing language has good support for the XML document object model (required if scripts must access an XML template). Templated datasets are considered part of a series, with tables having the same format and using the same metadata template. Approximately half of SBC packages (by number) are designed in this manner. Data are processed in the laboratory as appropriate, and a table output format is co-designed by the data owners and the information manager. The information manager creates an EML metadata template, which is filled in using semi-customized export scripts, by either by laboratory personnel or by the information manager. Using the same language as is used for data processing means that data publication can be more easily integrated with processing. An example of this pathway is our process for publishing stream discharge data using Matlab (Figure 3). A similar sequence is used for rainfall data and for sensor data from moored oceanographic instruments, also with Matlab. Standardizing the process has the advantage of furnishing a reliable product that can also be used as input for value-added products, e.g., a contribution to the network database, "ClimDB", in its required format (http://www.fsl.orst.edu/climhy). As of late 2013, metadata are now housed centrally, in a relational database system (see below). The IMS team is enhancing dataset production code to integrate template production with central metadata storage.

The second pathway ("unique", Figure 2, right) is appropriate for data packages in which table descriptions are unique (e.g., experiments), or where observations from distinct methodologies are kept separated (e.g., kelp forest community surveys). About half of SBC's packages are maintained using this pathway, and of those, about two-thirds are "ongoing time-series". This pathway is also necessary when the processing language does not have adequate support for the XML-DOM, or for which customized code for a single data package is not practical. In this pathway, laboratory personnel propose and assemble the unique content in text documents (e.g., title, abstract, table column descriptions), and the information manager reviews the proposed content, and adds it to SBC-Metabase. An assistant can handle SBC-Metabase insertion. We expect that as processing software progresses toward better XML-DOM support, that we may migrate some manually built data packages to a more automated pathway.

All data packages (both templated and unique) in the draft stage are presented to the scientists and/or staff using a "view draft" mechanism that displays metadata using the same views as are used by the production catalog. This enables scientists to better understand metadata components and approve the design before the package becomes final.

3. STATUS AND MANAGEMENT OF IMS COMPONENTS

3.1 Website

Essentially all SBC information is available via the Internet. The public website (http://sbc.lternet.edu) is organized around broad subject areas (Table 2) and complies with LTER Network standards for content, menus and links (LTER, 2009). The SBC website is a hybrid of static HTML, scripted pages and web applications. The unique content for static pages is maintained separately from the standardized content, menus and styling, and page assembly is scripted. As much as possible, material for links and menus is parameterized in configuration files. The modular nature of our website is consistent with general recommendations for web design, and also has advantages for SBC's co-management style; page content can be easily edited (or new pages created) by someone with very little training in web authorship and static sections can be easily replaced by scripts or web applications as these are developed. To facilitate multiple authors, all content is maintained in a version control system (Subversion). Responsibility for website coordination and integrity lies with the information manager. SBC also maintains two secure areas for SBC members after login: a) an internal website (https://sbc.lternet.edu) for non-public HTML-based material and b) HTTP-accessible data directories (https://sbc.lternet.edu/data). Details of the websites' design and implementation and documentation for individual applications are available in separate documents. Planned updates for the website are listed in the timeline (Appendix III).

Table 2. Major sections, content and implementation strategies for the SBC website (http://sbc.lternet.edu)

Section title	Content	Implementation
Research	Descriptions of core research activities	Backend storage for research themes is Metabase with export in projectDB. Browsed using Perl CGI script. (O'Brien, 2011).
Site	Descriptions and maps of sampling sites	Sampling sites are presented using Google maps. Considerable static content was deprecated in 2012.
People	Personnel directory and individual profile pages	PERL CGI application using LDAP database as the backend.
Publications	Bibliography, including presentations	Management in an XML-enabled RDBMS, with export EML Perl CGI to display and filter by author.
Data	SBC's data catalog, plus links to other catalogs of interest	The data catalog can be browsed using a Perl CGI script. Backend storage was migrated in 2015 (O'Brien, 2010)
Education and Outreach	Descriptions of K12 and higher education programs	Static content maintained by the project coordinator
News	Recent events of general interest, archived past events	Static content on the main page is updated quarterly by the project coordinator. News archive is a PERL CGI application with storage in text and RSS feed available (maintained by the information manager).
Information management	Information plan, guide, system and project documentation	Static content maintained by the information manager
Internal areas	Access to the fileserver, various help pages and site map	HTTPS, strong encryption. Certificate provided by LTER network.

3.2. Metadata

Many metadata components (e.g., personnel, methods, sampling sites) are necessary to describe data or to link between areas of the website. Metadata at SBC was historically managed ad hoc, but in 2013, we finished our migration into a relational database, with a centralized system making metadata more available for mulitple uses. Working with the Moorea Coral Reef (MCR) LTER, we adopted the database model that drives Georgia Coastal Ecosystems (GCE) LTER, "Metabase", and which also had been adopted by Coweeta (CWT) LTER. Initial adoption was to house descriptions of research themes with export as LTER-project XML (a subset of EML schema), which quickly demonstrated Metabase's usability and provide an entry point for work with Metabase design (O'Brien, 2011). In 2013, we finished code modules for export of EML datasets. These represent a significant improvement over the scripts originally written for project-export, and adhere to more professional programming practices. As this is a collaborative effort, all scripts and database work was coordinated between the two projects. Scripts were designed to work as web services, allowing them to facilitate synchronization with network databases in the future. Tasks for integrating Metabase with the SBC IMS are in the timeline (Appendix III). Because the Network Information System's PASTA framework (Provenance-Aware Synthesis Tracking Architecture, http://nis.lternet.edu) will require data packages with consistent, high-quality metadata, many of these activities are timed with that development in mind. Documentation of metadata systems can be found in the SBC IM Guide.

3.3 Data Packages

Data and metadata together comprise a "data package." Metadata documentation is published in the XML specification EML, exported from SBC-Metabase. ASCII tables are used for data delivery and archive, as these have proven to be the most flexible and efficient for long-term use of heterogeneous

data. Our goal as a data provider is to publish data packages in a state that is appropriate for their intended use, and we recognize that there is a spectrum of quality, completeness and complexity for both data and metadata. It has been our experience that for data to be confidently used in sophisticated applications, both high-quality data and metadata are essential. Simply posting data tables is not enough; data providers must consider additional features. SBC has developed definitions for data packages that describe their metadata and data content and quality (Table 3). Substantially more effort is required to produce the highest level ("Integration"), from which the data can be accessed by an application using only the EML metadata (e.g., a relational database).

All SBC EML metadata are contributed to the system maintained by LTER Network Office (LNO.lternet.edu); central to this is PASTA, which became the primary Network catalog in 2014 and houses both data and metadata. The LTER network integrated a quality engine into the PASTA data to ensure that the data it delivers are suitable for automated use. The list of criteria for acceptance by PASTA was accomplished by an LTER working group led by SBC information manager O'Brien in an LTER working group (Servilla et al, 2013; O'Brien, et al, in prep). As of this writing, there are 32 checks to be passed by every data package (with approximately 30 more to be implemented). All SBC data packages are contributed to PASTA, and all pass all appropriate checks. As the PASTA system is designed to deliver data for automated use, this means that essentially all SBC's data holdings are at the "integration" level. As automated use develops, other checking criteria may be introduced.

Data packages can be classified or sorted according to many other criteria, in addition to those above for metadata content and usability. For management purposes, SBC refers to these groupings as "inventory types", and several are incorporated in to our data package management software (see Section 3.3.1). The inventory in Appendix II is grouped according to some of these categories.

Table 3. Levels of data package complexity with potential uses. Asterisks indicate that related issues and needs are being considered at a network level. The conceptual "LTER metadata completeness levels" given in the last column are obsolescent, but are included for comparison.

Level	EML Metadata content	Data content	Uses
Information	Boilerplate: project description, publisher, contacts, access and use statements Unique: title, abstract, personnel*, publication date, Temporal, geographic and taxonomic coverage, keywords*	Data are optional	Information only, e.g., LTER Type II data LTER "Identification" or "Discovery" (see text) Searches by time, location, taxonomy Dataset citation
Download	**Information** + Data entity description, including column names, definitions & units*, physical description*, download URL	Data are required, but inconsistencies are allowed	Data are available, but the user may need help with interpretation. LTER "Evaluation", "Access", "Integration" (with caveats)
Integration	**Download** + Metadata congruent with data*	Data are required, and must be clean and congruent with metadata	1. Integration, workflows, further automated processing 2. Query applications 3. Contribute to Network databases LTER "Integration"

We are also planning for the incorporation of more advanced features that will further enable streamlined discovery and use, such as semantic annotation, unique and regularized identifiers, detailed methods and sampling design in metadata markup, and quality-controlled data values (Table 4). Such features do not necessarily require metadata or data of a certain complexity (as in Table 3), and potentially could be added to any data product. However, incorporating these features will require the development of more sophisticated tools, or are most effective to implement after significant standardization has occurred at a network level. Generally, it is expected that these features will not appear in SBC data products until after all the challenges of basic data ingestion/integration have been met. SBC participates in community and network projects that affect or promote these enhancements.

Table 4 Enhanced features of data and metadata. These metadata features can be incorporated at various levels, and data/metadata requirements vary. Most aspects of enhanced features involve community collaboration (asterisks).

Level	EML Metadata content	Data content	Benefits
Linked	Protocols Detailed project* descriptions	Any level, integration level is preferred	Links on website between data packages and other website areas, (research, publications)
Quality controlled	QC metadata* methods	Integration level, plus data values subjected to quality control measures*	Confident integration
Annotated	Semantic annotation*	Data optional, integration level is preferred	Confident and accurate discovery, automated integration

3.3.1 Data Package Management

In collaboration with MCR, software tools for managing these two sites' data package inventory were begun in 2012, and the RDBMS schema integrated with the Metabase schema (Gastil-Buhl and O'Brien, 2013). As with other Metabase work, SBC and MCR design and implement together, and then populate their individual instances. Tools are outlined in Table 4, and are being designed with the following use cases in mind: a) to share package-status information with an IMS assistant performing updates and maintenance, b) to sort/filter packages by inventory type, e.g., to generate an appendix for this document, and c) to display lists of packages on the web by status or type, e.g., packages in 'draft', or the inventory of data packages in Appendix II.

Table 5 Data inventory management tools in use at SBC.

Name	Description	Implementation	Status (current year)
Package management database tables	SQL tables for tracking status of data packages and their inventory types	postgreSQL	Tables fully populated for ongoing time-series packages
Web views	HTML views of data packages	PHP (from postreSQL)	dataset_status_tracking.php Linked from SBC website (internal, requires login).
SBC activity log	Links between sampling sites and research activities since the project's inception.	postgreSQL table with cross-references between research activities and sites	Populated time-series activities from watershed group, and reef group for 2000-2013, based on data holdings. Info to be checked and/or confirmed by laboratories.

3.4 Site Descriptions

All SBC time-series sampling sites have been described in an XML format that is compatible with EML

and KML. We are currently evaluating integration pathways for merging SBC's historical XML-based site descriptions with SBC-Metabase. Locations are displayed on the SBC website in two ways. First, each EML data package's HTML display shows a map as it is rendered by the data catalog (O'Brien, 2012). Secondly, our XML-formatted site descriptions are used to populate a Google map in the "Site" area of the SBC website, to provide spatial context for research (Table 2). Geographic information will also be made available as KML, which is appropriate for many external applications including mapping tools being developed by the LTER Network (Appendix III). We continue to review policies and practices for managing SBC sampling sites (https://sbc.lternet.edu/info_management/research_sites.html).

3.5 Personnel

The current backend storage for our personnel database also houses the system for fileserver access (LDAP), and is maintained by MSI personnel and co-managed by MSI, MCR LTER and SBC. Personnel information from LDAP was ported to Metabase during 2011 as part of the adoption of Metabase for research projects. Personnel information for data packages is exported from Metabase (not LDAP), and we anticipate that during the second half of SBC III, the web personnel directory will be served by Metabase content, and SBC's LDAP will be used only for fileserver access (timed with other MSI upgrades, Appendix III). Until then, personnel information must be maintained in both LDAP and Metabase, and also in the Network personnelDB. Ideally, site personnel information will be synchronized with the Network database via web services, which we anticipate to be feasible during the second half of SBC-III, and pending refunding of the current LNO (Appendix III).

3.6 Research Activities Catalog

High-level SBC research themes were described and added to Metabase in 2011. These are exported for web display in the LTER-projectDB XML specification and are linked to the personnel directory. The LTER-project schema is based on EML, and is in use at several LTER sites. To facilitate linkages between SBC research themes and relevant data, the project catalog uses the same menus and keywords as those used for datasets. We plan to expand our use of projectDB to include details of specific research projects and sampling activities, which will enable more website cross-links between research, data, sampling sites and/or publications (Appendix III).

Knowledge of ongoing research activities is also applicable to data package updates and management, e.g., so that IM personnel can anticipate new data. In 2012, initial content was added to Metabase for ongoing research activities and sampling sites, and these two tables cross-referenced in the package management schema (Metabase, section 3.3.1, above). Ongoing time-series are planned to be the next group of SBC research activities to be fully populated and displayed with projectDB.

3.7. Bibliography

A research group's bibliography is an important publicly available resource. SBC adds 50-75 new citations annually which also must be managed in the databases of funding agencies (e.g., NSF's "report.gov") and at the LTER Network. In 2005, SBC moved its bibliography from a static text list to descriptions written in EML and housed in Metacat (O'Brien, 2006). In the process we contributed significantly to the development of both EML and Metacat. We have found the EML specification to be well suited to a bibliography, and maintained this feature of the bibliography as it was ported to the SBC-Metabase in 2014. Currently, citations are managed in SBC-Metabase, with citations as EML in XML. These improvements enabled reporting to NSF to be managed by the IMS, instead of ad hoc by

the project coordinator, and streamlined contributions via exports in BibTeX. Future plans include cross-linkages between personnel (as upgrades in that area progress), and to datasets (Appendix III).

3.8. Quality control and Protocols

Quality control is generally based in the researchers' laboratories. All data packages include methods, usually in the form of protocol documents (PDF). Data collection and processing protocols are easily accessible on the fileserver to SBC personnel, and a metadata system for a protocol bibliography available in Metabase. To date, we have outlined current practices and recommendations for managing protocol documents in data packages (https://sbc.lternet.edu/info_management/research_protocols.html). The information manager works closely with analysis personnel to document quality control in metadata as appropriate for individual data packages. Quality control of SBC's community survey data has been documented (O'Brien and Harrer, 2008). Documenting quality control at the data package level is an enhanced feature of data packages (Table 4), and planned for the second half of SBC III (Appendix III).

3.9. Standardized Measurement Descriptions

In order for successful synthesis to occur at a Network level, LTER sites will need to have described their measurements in such a way that these can be compared using automated tools, and/or registered their measurements with network data synthesis research projects. The most straightforward strategy is to first standardize measurements at the site level with complete descriptions (including methods, precision and error). Because this is an enhanced feature of data (Table 4), we anticipate this activity occurring during the second half of SBC-III. We will work with the LTER IM community as we standardize our measurements in preparation of this becoming a LTER Network wide requirement (O'Brien, 2010a). SBC is uniquely positioned to assume a leadership role due to our involvement with ontology development with the Extensible Observational Ontology (OBOE) in the Semtools project (DBI-0743429, O'Brien co-PI), and with the DataONE Semantics and Provenance working group (O'Brien, 2014).

3.10 Data Catalog

SBC's data catalog has been based on EML metadata since 2003. The catalog was decoupled from a local, legacy Metacat system in 2010 to draw SBC EML directly from the LTER Network for display on the SBC website (http://sbc.lternet.edu/data). SBC data packages are categorized into "collections" accessible via web forms keyed to local habitats, measurement types and LTER core research topics. As part of the migration in 2010, EML-dataset XSL templates were redesigned in a "tabbed" format to highlight major sections such as geographic area and methods. The SBC templates are used by two other LTER sites (Moorea Coral Reef, Virginia Coast Reserve), the Network Controlled Vocabulary website (http://vocab.lternet.edu), and contributed to the redesign of the Network-wide data catalog (http://portal.lternet.edu).

There have been incremental changes to the data catalog since 2010, most notably, the addition of templates for dataset citation (2011), a Google-map showing geographic coverage (2012), and addition of DOIs to citations (2015). Since SBC's catalog relies on the Network NIS for EML content, we recoded to use local stores when the Network NIS removes Metacat from its software stack in 2015. We amy consider other options for our catalog search and delivery (e.g., DataONE), as these may provide the ability to create queries for other data resources in the Santa Barbara area, not just SBC's. Work will be carried out with other LTER sites interested in the same features.

4. SBC Informatics projects

4.1 LTER Network

SBC information manager O'Brien co-chaired LTER Information Managers' Committee (IMC, http://im.lternet.edu, http://intranet.lternet.edu/committees/information-management) from 2009-2013. During active PASTA development, she was a member of two NIS Tiger Teams ("Data Manager" and "Metadata Quality") and was on the LTER Network Synthesis Data Committee (http://intranet.lternet.edu/committees/synthesis-data).

SBC's contributions to the Network are chiefly concerned with the quality and usability of EML metadata. O'Brien has served on numerous IMC working groups chairs, "EML Congruence Checker," and "Dataset Design." This leadership stems from SBC's early adoption of EML for our own catalog, and our experience with the EML Data Manager Library (section 3.11, above). Additionally, our use of EML in the SBC LTER bibliography was reviewed by the EML development community and contributed significantly to the enhancements to the EML schema version 2.1. O'Brien served as the EML 2.1 release coordinator (O'Brien and Jones, 2008). Our work with LTER-project XML, based on EML and co-led by O'Brien and C. Gries (North Temperate Lakes LTER), is also likely to contribute to EML advancement. O'Brien previously chaired the "EML Best Practices" working group (2009), and also serves on two other IMC working groups, "UnitsDB" and "Controlled Vocabulary". All three are concerned with the standardization of EML dataset content. SBC plans to use the web services of the latter two systems as they become available (and are maintained), most likely after 2015 (Appendix III).

4.2. OBOE Ontology

Our work with the LTER Controlled Vocabulary is also related to O'Brien's efforts with ontology development with the Extensible Observational Ontology (OBOE) in the Semtools project (DBI-0743429, Leinfelder et al. 2011). This work also has the capacity to inform similar ontology development at the Network level, for example, in data discovery or the description of standardized measurements, and will also facilitate interoperability with systems beyond the LTER Network, such as the Biological and Chemical Oceanography Data Management Office (BCO-DMO), and the Consortium of Universities for the Advancement of Hydrologic Science, Inc. Hydrologic Information System (CUAHSI HIS). SBC plans to examine the usability of the OBOE ontology for standardizing SBC measurements during the second half of SBC III (Appendix III).

4.3. DataONE/GeoLink

O'Brien was invited to join the DataONE working group "Semantics and Integration" in 2013. DataONE is a data federation project funded by NSF-ACI (http://dataone.org). LTER is a "primary DataONE node", and consequently, all LTER data appear there. O'Brien's involvement is concerned mainly with describing the context and use cases for scientific data. In 2014, she formally joined the DataONE project to develop improved search mechanism using LTER primary production data. O'Brien is also involved with the GeoLink project, funded by EarthCube. This effort compares schemas and models common information to enable cross-schema searches of data from disparate contributors such as LTER and the database maintained by the Biological and Chemical Oceanography Data Management Office (BCO-DMO.org).

4.4 Data Query for EML packages

Because tables associated with time series can become large and cumbersome, we developed a generic tool (the EML Data Query tool, EDQ) for loading Integration-level datasets into a relational database so that data can be queried via web forms (Figure 4, O'Brien and Burt 2007, (Leinfelder et al 2010). The application is not customized to any single dataset type. It reads EML metadata, uploads the described data table to a relational database and creates a map interface and form that then generate SQL queries based on user input. The application takes advantage of established community standards and accommodates a variety of data tables. This approach allows data owners to control the format of the tables they publish, while accommodating a repository of highly varied scientific data, and still allows the complete table to be archived in ASCII format. Another alternative would be to create custom interfaces and data models for each data type; however, that added complexity would increase maintenance costs and further strain resources.

The EDQ was written in 2006 for EML 2.0.1 and uses a prototype of a Java library written by the LTER and the NCEAS Ecoinformatics programming groups. This code library (Data Manager Library, or DML) is now being significantly revised as a component of the quality engine in the LTER NIS PASTA framework. A newer version of EML (2.1) has significant advantages over 2.0.1, and all SBC datasets were be upgraded to meet its requirements. Consequently, the EDQ must be redesigned or replaced. We had tentatively planned this activity pending certain other activities not under our control, mainly the NIS production release and a revised DML. However, the subsetting of large datasets is a community need, and so no single research group (such as SBC) should develop a custom, local solution. Since SBC has substantial experience in this area, it's more likely that we will be involved in an advisory capacity within a much larger group such as the LTER Network or even DataONE, both of whom have expressed an interest in their systems meeting this need.

A need to subset several large SBC datasets remains. In recent years, as an additional quality control measure, we now regularly ingest data tables into a relational database system mainly to check data typing and to compute data ranges and explore basic features (e.g., descriptive statistics). An added benefit of this practice is that once ingested, data could also be subset manually using SQL queries (by O'Brien). This capability has not been widely advertised, and such a service would be custom and ad hoc. However, it remains as an alternative if other avenues are not fruitful.

Literature Cited

Gastil-Buhl, M. and M. O'Brien. 2013. Data Package Inventory Tracking: Slicing and Dicing with SQL. LTER Databits, Spring 2013. http://databits.lternet.edu/spring-2013/data-package-inventory-tracking-slicing-and-dicing-sql

Leinfelder, B., S. Bowers, M. O'Brien, M. B. Jones, M. Schildhauer. 2011. Using Semantic Metadata for Discovery and Integration of Heterogeneous Ecological Data. Proceedings of the Environmental Information Management Conference. Santa Barbara, September 2011.

Leinfelder, B., J. Tao, D. Costa, M. B. Jones, M. Servilla, M. O'Brien, C. Burt. 2010. A metadata-driven approach to loading and querying heterogeneous scientific data. Ecological Informatics, 5: 3-8.

LTER 2009. Review Criteria for LTER Information Management Systems.

LTER 2009a. Guidelines for LTER Web Site Design and Content.

LTER 2010. Strategic and Implementation Plan. http://intranet.lternet.edu/documents/lter-strategic-and-implementation-plan

O'Brien, M. 2014. DataONE to enable semantic searches for LTER NPP data. LTER Databits, Fall 2014. http://databits.lternet.edu/fall-2014/dataone-enable-semantic-searches-lter-npp-data

O'Brien, M., M. Servilla and D. Costa. In prep. Assuring the quality of data packages contributed to the LTER Network Information System. to be submitted to Ecological Informatics, special issue, Dec 2015.

O'Brien M. 2012. EML and Google Maps. LTER Databits, Fall 2012. http://databits.lternet.edu/fall-2012/eml-and-google-maps

O'Brien M. 2011. The Santa Barbara Coastal (SBC) LTER's implementation of projectDB using Metabase. LTER Databits, Fall 2011. http://databits.lternet.edu/fall-2011/santa-barbara-coastal-sbc-lters-implementation-projectdb-using-metabase

O'Brien M. 2010. Using EML in Your Local Data Catalog. LTER Databits, Spring 2010. http://databits.lternet.edu/spring-2010/using-eml-your-local-data-catalog

O'Brien M. 2010a. Using the OBOE Ontology to Describe Dataset Attributes. LTER Databits, Fall 2010. http://databits.lternet.edu/fall-2010/using-oboe-ontology-describe-dataset-attributes

O'Brien M. and C. Burt, 2007. A Query Interface for EML dataTables. LTER Databits, Spring 2007. http://databits.lternet.edu/spring-2007/query-interface-eml-datatables.

O'Brien M. and S. Harrer. 2008. Processing and quality control of kelp forest community survey data. Proceedings of the Environmental Information Management Conference, Albuquerque, September 2008.

O'Brien M. and M. Jones. 2008. EML 2.1.0 to be Released Soon. LTER Databits, Spring 2008. http://databits.lternet.edu/spring-2008/eml-210-be-released-soon

O'Brien, M. C., C. Gries, W. Sheldon, J. Walsh, J. Porter, S. Bohm, J. Brunt, S. Remillard, K. Ramsey, J. Downing, K. Vanderbilt, R. Aguilar and M. Servilla. 2009. A collaborative database for LTER projects at sites. Poster at: 2009 LTER All Scientists Meeting, Estes Park, CO.

SBC LTER. 2009. Attributes and Units in LTER Data Packages, v0.9. http://sbc.lternet.edu/external/InformationManagement/documents/SBC/Attributes_Units_LTER_data_packages.pdf

Servilla, M., M. O'Brien and D. Costa. 2013. Assuring the quality of data packages contributed to the LTER Network Information System. American Geophysical Union Fall 2013 Meeting, Poster IN53C-1576.

Literature no longer cited in this document, but pertinent and indicative of SBC IMS development:

O'Brien, M. C. 2006. Using EML and Metacat for a site bibliography. at: 2006 LTER All Scientists Meeting, Estes Park, Colorado.

SBC LTER. 2008. Guide to Creating SBC Datasets with Morpho, v0.9. http://sbc.lternet.edu/external/InformationManagement/documents/SBC/Morpho_userGuideSBC .pdf

Appendix I. SBC Hardware infrastructure

Computing systems at MSI

Data management for the project has the advantage of utilizing the computing capabilities of the Marine Science Institute (MSI). MSI has a 1000Mb/s connection to the UCSB campus backbone, which provides shared access to a 622Mb/s CALREN-2 connection, which in turn provides access to the Internet.

MSI supports the research servers. The main data server providing network file sharing (Samba and NFS) is a running CentOS Linux 5 (64-bit), with a web server running CentOS Linux 6 (64-bit) running Apache. The primary database server for Research use is running CentOS 6 (64-bit) providing PostgreSQL 9.2, and MySQL 5.1. There is a departmental server available for Subversion and Git revision control systems as well as a computational server running CentOS Linux 6 (64-bit) with Matlab, R, GSLIB and PERL along with other scientific applications. MSI runs an Enterprise-grade SAN for storage. Currently we have 12 TB of storage allocated (and expandable) available on that system. The Server room is connected to E-Power, and redundant power provided by an APC 6000 UPS battery backup. Distributed server backups (via Amanda and BackupPC) are coordinated with MSI.

File system backup

Backups are carried out by the MSI computing staff. Full backups (level 0) are performed monthly, with incremental (level 5) and progressive incremental (level 9) backups weekly and daily, respectively. Five months of disk-to-disk backups are stored on the server, with storage space allocated to the /backup partition as necessary.

Appendix II. SBC Data Inventory

January 2013

The inventory of data packages is presented in two parts: Appendix II.1, Cataloged data package, and Appendix II.2, Data packages currently in draft, undergoing re-design, or for which data are not yet released but expected. The inventory is online as of early 2013; please note date on each page. Information for both types is likely to change.

Appendix II.1. Inventory of publicly available SBC data packages

Data packages are classified by SBC-III proposal type, use/intent code, management type, and network type. EML version, date of most recent update and number of data entities are included. This inventory is organized by management needs. Packages are organized according to temporal types: 'ongoing time-series', 'short-term data', 'terminated time-series', and 'non temporal' (generally, supplementary material). 'Terminated time-series' is not currently used, but may be reinstated.

Appendix II.1: Cataloged data packages:
https://sbc.lternet.edu/internal/research/Metadata/EML_data_packages/00_pkg_management/inventories/cataloged.html

Appendix II.2. Inventory of draft, anticipated or Network Type II data packages

Data packages whose metadata are not yet pubic, under-described packages, or packages that do not yet have publically available data (i.e., will be moved from Network Type II to Network Type I) are listed here. Type II packages are likely to have had metadata already released (cataloged), and data manually distributed with permission from the PI. Data are available for to those with an SBC login. ALL INFORMATION HERE IS PROVISIONAL.

Appendix II.2: Draft and anticipated data packages:
https://sbc.lternet.edu/internal/research/Metadata/EML_data_packages/00_pkg_management/inventories/draft_anticipated.html

Appendix III. Time line of SBC IMS Improvements

(Last update: January 2015)

Table III.1. IMS improvements as proposed in SBC-III Proposal Supplement 2-4, reproduced in table form.

item	*year*	*Proposed plans*	*Deliverable*	*Progress*
1	2011-2012	All SBC time-series sampling sites have been described in standard format and applications are in development to map these on the website to provide spatial context for research and data. Geographic information will be available as KML, which is appropriate for many external applications (e.g. mapping tools being developed by the LTER Network).	Mapped sampling sites	Done. Maps from both /sites/sampling, and in EML dataset templates
2	2012	Re-build all EML datasets from SBC-Metabase. We began building datasets with Metabase in 2011. We plan to update our time-series datasets in 2012 from Metabase, and upgrade all static datasets to EML 2.1	Higher quality metadata	Done. 2012-2013, for datasets existing in 2013. Also developed code and workflows for design and maintenance
3	2012	The development version of the NIS Provenance-Aware Synthesis Tracking Architecture (PASTA) is scheduled to be available in mid-2012. We will submit all of our datasets to the NIS	SBC data tested against PASTA system	All datasets submitted regularly for evaluation since 2012.
4	2014	PASTA is scheduled for production in 2014. We will submit all of our datasets at that time (dependent on NIS software meeting its own milestones) Note: In 2012, NSF accelerated the production schedule; release set for Jan 2013.	SBC data in PASTA (by the end of 2013)	Done. SBC datasets contributed Dec 2013, with updates and new packages ongoing.
5	2015	Expand the content of SBC's projectDB to include details of specific research projects and sampling activities	Website cross-links between research, data, sampling sites and/or publications	Began in 2012, see table 2 for details.
6	2016	In order for successful synthesis to occur at a Network level, sites will need to have described their measurements in such a way that these can be compared using automated tools, and/or registered their measurements with network data synthesis research projects. The most straightforward method is to first standardize measurements at the site level with complete descriptions (including methods, precision and error). We will work with the LNO and the LTER IM community as we standardize our measurements in anticipation of this becoming a LTER Network wide requirement.	Standardized SBC attributes	Ad hoc regularization as part of a) Metabase population in 2013 and b) continued dataset update and design in 2014

Table III.2. Following are the specific activities required to produce IM System deliverables in Table III.1. Proposed deliverables from Supplement 2 are in **boldface** and the line numbers of major deliverables (Table III.1) are referenced. Some activities relate to more than one deliverable, or relate to all by improving efficiency. The timeline focuses on the period before the mid-term review in 2015. Column labeled "Affected IMS component" contains a reference to the IM Plan section. If no individual's name is given, the site information manager and/or assistant (as appropriate) will perform the activity.

Not included, but ongoing in all years:

a) Design and cataloging of new data packages, updates of existing data packages (code-creation is noted)
b) Server and software upgrades and migration
c) Participation in Network-level projects, except as noted

Year	*III.1 item*	*Deliverable*	*Affected IMS component*	*Activity Description*	*Progress as of Jan 2015*
2012	1	**Mapped sampling sites**	Website, site descriptions (3.1, 3.4)	Finish web scripts to map SBC time-series sampling sites on the website to provide spatial context for research and data. Make geographic information available as KML	Done
2012	2	Metabase populated from existing datasets	Metadata (3.2)	From various sources (existing EML, LDAP, manual insertion) populate the Metabase tables necessary to export EML datasets	Done in 2013
2012	2, 6	Consistent look-up tables	Metadata, data packages (3.2, 3.3)	Regularize values of keys by converging on consistent content in lookup tables. Via queries to Metabase, analyze components for inconsistencies. Includes: personnel, data package titles, abstracts and keywords, entity-level names, physical descriptions, abstracts, delivery URLs (e.g., Network Data Access System, DAS)	Done in 2013
2012	2	Efficient export of EML from Metabase (**Higher quality metadata)**	Data packages (3.3)	Expand scripts initiated in 2011 for EML export from Metabase to build complete datasets. All datasets will be exported as EML 2.1.	Done, 2013 (P.Slaughter, O'Brien)
2012	3	**SBC data tested against PASTA system**	Network Information System (NIS)	Submit SBC datasets to the PASTA-prototype system scheduled to be available in 2012	Ongoing; all datasets tested regularly during draft or update.
2012	All	GCE Matlab toolbox workshop participation	Planning, data packages (3.3)	GCE LTER has developed Matlab tools that streamline some aspects of dataset description and manipulation, and which work with Metabase. A workshop is planned to introduce the toolbox to other LTER sites.	Done, November 2012 (O'Brien, Gotschalk)
2013	5	Website upgrades	Website (3.1)	Redesign some static sections. Integrate new menus. Incremental changes to dataset and project views. Update research-theme content for SBCIII	2012, some menu components available (M. Thompson). Reschedule after webserver migration, 2014

Year	*III.1 item*	*Deliverable*	*Affected IMS component*	*Activity Description*	*Progress as of Jan 2015*
2013	All	GCE Matlab toolbox evaluated for SBC use	Planning, data packages (3.3)	Evaluate the usability of GCE Matlab toolbox for SBC, with respect to a) our implementation of Metabase, b) SBC's existing data package publication workflows in Matlab, and c) production of value-added datasets from workflows. Collaborative with MCR.	In progress, by Chris Gostchalk (supp. funds) and O'Brien
2013	5	Outline possible pathways for EDQ upgrade	Planning, website, data package management , EDQ (3.1, 3.3, 3.11)	The EML dataset query application (EDQ) was written in 2006 for EML 2.0.1 using a prototype of the Data Manager Library (DML, Java code). It requires upgrade and/or redesign. Both EML and the DML have been significantly revised (e.g., EML 2.1, and the DML as a PASTA component). The upgrade scenarios, or plans for replacing this functionality will be outlined.	Done. This is a broad need. Solution should not be site-specific. SBC will contribute use cases to Network or DataONE.
2013	2, 3	SBC/MCR web service code for network DB synchronization	Data packages (3.3)	SBC has supplement funds to refine software for export of EML datasets from Metabase. Work includes a "switchboard" or some equivalent mechanism for synchronization with Network DBs via web services. This activity will require collaboration with MCR, and hiring of a part-time or short-term enterprise level programmer. As of 2012, web services are available centrally from the Network unitsDB and vocabDB (keywords). As of 2014, no other web services are complete, nor is the future of any guaranteed.	2013, proof-of-concept code completed (for http forms , P. Slaughter, supp. funds). Continued work on hold until Network DB's stabilized and fully implemented
2013	2, 3	Semi-automated data update workflows for ongoing datasets	Data packages (2.2, 3.3)	Refine existing scripts (Perl or Matlab, TBA) or adopt new tools (e.g., GCE toolbox, Metabase) for semi-automated updates of ongoing datasets. a. For "unique" pathway: SQL-XML queries and new code. b. For "template" pathway, integration of SQL-XML into existing code	a. Code done 2013 (P. Slaughter, supp funds); SQL-XML queries (O'Brien) b. 2014, upgrade code repository, conceptual layout (O'Brien, Gotschalk)
2014	5	Upgraded EDQ	Website, EDQ (3.1, 3.11)	The EDQ will be upgraded or its functionality replaced **by LTER Network or DataONE**	N.A. see text.
2014	3, 4	All SBC datasets are EML 2.1 or better	Data packages (3.3)	SBC datasets that have remained at EML 2.0.1 to be compatible with the EDQ must be upgraded because PASTA will not accept EML 2.0.1.	Done, 2013
2014	3, 4	**SBC Data in PASTA**	NIS	Submit all SBC datasets to the production PASTA system scheduled for release in 2014. Schedule was accelerated in 2012. See text.	Done, 2013

Year	*III.1 item*	*Deliverable*	*Affected IMS component*	*Activity Description*	*Progress as of Jan 2015*
2014	5	Lower technical bar to maintaining project information	Metadata (3.2, 3.6)	ProjectDB was populated in 2011 with research project themes by the lead IM. We would like non-IM personnel to assume this task, which require web forms for input to Metabase. In 2013, SQL-XML queries were developed for datasets (for this purpose), and are being tested/hardened in 2014. These could serve as the model for project maintenance.	Likely postponed. Supplement funds are not anticipated so work will rely on primary info. mgr. time.
2014	5	SBC project themes on the web are linked to activities	Website, research projects (3.1, 3.6)	Expand use of projectDB schema and associated scripts/templates for displaying scientific activities in Metabase on the website.	2012: began populating ongoing time-series activities in metabase
2014	5	Metabase evaluated for SBC bibliography	Bibliography (3.7)	SBC's bibliography is currently stored in extensive EML, with additional metadata. Possibly, management could be more efficient with different storage. Examine Metabase for this use, particularly with regard to required reporting and linkages on the SBC website.	Evaluation done 2013; see below for execution.
2014	5	Import SBC bibliography to Metabase	Bibliography (3.7)	Add biblio schema. Use GCE and MCR implementations as model for SBC. Maintain citations in XML (using postgres XML type); add maintenance fields.	2014: done
2014	5	Redirect search and display functions of SBC data catalog away from LNO Metacat	Website, Data Catalog (3.10)	SBC's data catalog 'pulls' dataset content (EML) from the Network data catalog, Metacat. As part of the LTER NIS development, LNO plans to discontinue use of Metacat. SBC's dataset search and retrieval functions must be recoded to use a) the NIS PASTA system, b) DataONE, or c) our local Metabase. (Collaborate with MCR)	Conceptual layout. evaluation of options and SOLR queries. LNO Metacat becomes unavailable in 2015
2015	5	Migrate from LDAP as personnel database to Metabase	Website (3.1)	Recode personnel directory and individuals' profiles to pull content from Metabase instead of LDAP	
2015	5	Cross-linked Metabase tables	Metadata (3.2)	Complete any additional cross-links between research, data, sampling sites and publications in Metabase. Cross-links will have been initiated as part of earlier work with data set production.	
2015	5	**SBC website cross-links between research, data, sampling sites and/or publications**	Website (3.1, 3.2, 3.3, 3.5, 3.6, 3.10)	Expand and harden website scripts, and standardize configuration to streamline display and maintenance of cross-linked information in Metabase. Specific activities TBD.	

Year	*III.1 item*	*Deliverable*	*Affected IMS component*	*Activity Description*	*Progress as of Jan 2015*
2015		Mid-term review	All	Work-to-date will be highlighted for the mid-term review (October, 2015)	
2015		Exploit centrally-stored metadata	Data packages (3.3, 3.5)	It is expected that by 2015, additional production NIS modules will supply content for EML datasets via web services (personnelDB, biblioDB). Also, all central database are also planned to synchronize with site systems via web services.	Dropped. Improvements to LNOl DBs are no longer anticipated
2015	6	Evaluate community measurement ontologies	Planning, standardized measurement descriptions (3.9)	Plan for use of existing knowledge models for SBC's measurement standardization needs (e.g., hierarchical vocabularies, ontology). This work follows from earlier work with SBC datasets and the OBOE ontology (2009-2014).	2014: collaborate with DataONE to design measurement-level ontology for LTER-wide primary production data
2016	6	**Standardized SBC attributes**	Data packages, NIS (3.9)	Standardize SBC measurements with complete descriptions, including analysis methods, precision and error.	2014: Regularize SBC variables associated with kelp forest community survey data and ocean acidification collaborations
2017		SBC-IV proposal IM section	All		

Appendix IV. Research collaborators and relationships with the SBC LTER IMS

(Last update: January 2015)

Table 1. The following projects collaborate with SBC LTER, and their relationships with the SBC Information Management System (IMS) vary. There are several patterns for the extent of the involvement of the SBC IMS; these examples are not exhaustive, and some projects fall into more than one:

a) The SBC IMS has no involvement with the project's data management.

b) Data are housed on SBC server, and arrangements are informal. No plans exist for publication by SBC. This is likely to be true for projects that predate NSF's required data management plan for all proposals (2011).

c) Data are housed on SBC server, and SBC plans to publish data; priority determined ad hoc.

d) Project has a data management plan that leverages SBC IMS and will publish data through our catalog.

Years active	*Project*	*Funding*	*PIs*	*Data description*	*Extent of management by SBC IMS*
19xx-	CODAR	varies	Washburn	Surface currents from radar	None. CODAR maintains it's own data catalog
2005 - 2007	CEQI (Coastal Environmental Quality Initiative)	UCOP (UC Marine Council)	?Reed, ?Gaylord, Stewart?	Moored instruments co-located at SBC reefs	On file server. SBC will publish data. also see: http://escholarship.org/uc/search?entity=ucmarine_ceqi
19xx -	PISCO (http://piscoweb.org)	Moore, Packard	Gaines, Warner	Sampling area overlap SBC near Pt. Conception	None. PISCO maintains it's own data catalog.
1994-	Plumes and Blooms (P&B) http://pnb.eri.ucsb.edu	NASA	Siegel	Semi-monthly profiles of CTD and optics across the Santa Barbara Channel	P&B maintains its own data catalog. SBC will develop data products from CTD profiles with cooperative plan for future inclusions. Timeline tbd.
20xx - 2014	CALobster		Lenihan	Fisheries surveys	
	CDIP (SIO)	MMS		Modeled ocean swell and circulation in the Southern Cal bight	Model results for selected SBC reefs are retrieved nightly and republished by SBC IMS (Type 0 data)
	Channel Islands National Park	NPS		reef community surveys	None
2012 - 2015	Channel Islands National Park	NPS	Kapsenberg	nearshore pH	Data will be managed and published by SBC
	Channel Islands National Marine Sanctuary				None
	Santa Barbara Land Trust (NGO)		Melack	Water chemistry data from Arroyo Hondo	Water chemistry analyzed and published with other SBC data
	Santa Barbara Channelkeepers (NGO)		Melack	Water chemistry data from Ventura River	Water chemistry analyzed and published with other SBC data

Years active	*Project*	*Funding*	*PIs*	*Data description*	*Extent of management by SBC IMS*
	City of Santa Barbara	City of SB	Melack	Water chemistry in Mission and Arroyo Burro creek catchments	Water chemistry analyzed and published with other SBC data
	County of Santa Barbara	County of SB	Melack	Water chemistry	Water chemistry analyzed and published with other SBC data
	Friends of the Santa Clara River (NGO)				
	USGS			Stream discharge	SBC publishes processed data from USGS gauges in our area (Type 0 data)
2010 -	Coastal-trapped waves	NSF	Fewings	Physical oceanography	Occasionally housed/processed on SBC/MSI servers. No plans for SBC to publish data
2010 -	Regional upwelling relaxation	NSF	Washburn, Fewings	Physical oceanography	Occasionally housed/processed on SBC/MSI servers. No plans for SBC to publish data
2010 –	Reef foodweb	NSF	Page, Miller	C&N Isotope content of producers and consumers	SBC will publish data
2005 – 2007	Kelp Biomechanics	NSF	Gaylord		On file server.
2012 –	Ambient pH	State of California	Passow	Bi-weekly pH of seawater for instrument calibration	SBC publishes data along with it's own that is similarly collected and processed
2010 –	AUVs		Siegel	CTD deployments on autonomous vehicles	SBC will publish data
	Surf grass		Reed, Blanchette		On file server.
2004 -?	Kelp Dispersal		Reed		On file server.
	Regional reef community		Rassweiler		On file server.
2010 –	Diatom exudates and high CO2	NSF OCE	Passow, Brzezinski, Carlson		None. Project uses BCO-DMO
2012 –	Kelp genetics	NSF	Alberto, Reed		SBC will publish data
2012 – (2015)	OMEGAS	NSF 1220359	Hofmann, Blanchette, Washburn	ambient pH, carbonate system params	SBC does not manage data. Collaborating on processing and publication methods. Project to use BCO-DMO.
2013	CCE's SBB channel (project started in 1993)	CCE/SBC	Benitez-Nelson, Thunell	sediment trap time series	TBD; possibly a shared dataset. PIs primary site is CCE, SBC maintains deployment

Years active	*Project*	*Funding*	*PIs*	*Data description*	*Extent of management by SBC IMS*
2014 -	Seastar wasting		Blanchette	ad hoc observations, pertaining to occurrence of disease	SBC plans to develop a data product on obs of the disease from its own surveys
TBA	PROPOSED: toxoplasmosis	proposed to OCE	Passow	toxoplasmosis in local water.	Probably we will host data.
2014 -	Dynamic and material connectivity	OCE 1355970?	McWilliams, (Romero?)	Diverse modeled data processed at ERI.	Unknown: metadata only (for linkage)? Primary archive TBD. Perhaps exogenous/republished
2014 -	Marine BON: diversity indices	NOAA, NASA	Miller, Rassweiler et al	integration of original and exogenous data to produce new indices	clone components of SBC IMS; O'Brien to lead (with .25 FTE)
2015 – (2019)	Kelp forest/sandy beach linkages	NSF OCE 1458845	Dugan, Ohlmann, Miller	drifters, intertidal community structure, consumption, ROMS modeling	SBC IMS will publish data