



# LTER Databits

Information Management Newsletter of the Long Term Ecological Research Network

01001100 01010100 01000101 01010010

## Fall 2012

In the Fall 2012 DataBits issue we feature articles related to sensor networks as a follow-up to the SensorNIS workshop held at Hubbard Brook Experimental Forest LTER in New Hampshire on October 25-27, 2011. We use this opportunity to present articles that discuss 1) progress made on sensor site establishment including types of sensors, sensor selection, or sensor platforms, 2) means of data transmission or communication networks, 3) development or adoption of data processing middleware, 4) quality control or data qualifying procedures, 5) data archiving methods, or 6) sensor management tools or solutions.

For one component of this issue we asked attendees of the SensorNIS meeting at either Hubbard Brook or the follow-up meeting at the LNO to offer short updates about how their sites had changed over the past year. We had hoped to have a short report from every site about their sensor-related activities, but instead we had several summaries and a plea for further development of management tools. Specifically, there were requests for standards, packaged tools, working demos or tutorials, and practice data. This issue comes at an opportune time to advance the discussion on sensor networks and, hopefully, further some of these standards that we set out to achieve over the past year. Please use this edition of DataBits as a springboard to keep the conversation going and work towards those goals.

Adam Skibbe (KNZ) [askibbe@ksu.edu](mailto:askibbe@ksu.edu) and Aaron Stephenson (NTL) [aistephenson@wisc.edu](mailto:aistephenson@wisc.edu)  
Fall 2012 Co-Editors

## Featured Articles

---

Information managers focus on data availability and use  
Sensor Network at the Andrews Forest  
Sensors at North Temperate Lakes  
Sensor Activities on the Virginia Coast Reserve

## Commentary

---

Targeted searches with EML and LTER Controlled Vocabulary  
A Request for Mr. S. Claus - Some Sensor Tools!

## News Bits

---

Sensors and Superstorm Sandy  
SensorNIS: Community engagement to build a resource guide for managing sensor networks and data  
EML and Google Maps

## Good Tools And Programs

---

A Web Service for EML-based Mapping

## Good Reads

---

IT Support and GIS/Remote Sensing Analyses are Vital for Disaster Rapid Response at UNOSAT  
Data Reporting and Data Usage, European Style

## Featured Articles

---

## Information managers focus on data availability and use

edit

Margaret O'Brien (SBC), John Chamblee (CWT)

### 2012 All Scientists Meeting

The Information Management Committee (IMC) meeting on September 9 at Estes Park was attended by about 40 site representatives and guests, and as usual, was the event at which we assessed our progress and prepare for the coming year. Activities focused on sites' data contributions to the network catalog, particularly the PASTA-related tools for gauging the structural quality of data and metadata, and key features of EML metadata. During the larger ASM, information managers organized several IM-related working groups and were deeply involved with other synthesis groups. Informatics-related posters were contributed from virtually every site and encompassed all aspects of data management. See the Network News for more information about the ASM, including IMC activities.

### New Working Group

To begin network-wide implementation of new PASTA-related tools to gauge dataset structural quality, the IMC formed a new working group tasked with designing reports for various stakeholders, e.g., sites, scientists, the Executive Board (EB), and NSF. The new working group has already met to outline the scope of reports, their timing and frequency, and relationship to the PASTA development timetable. The group is also considering ways to assess the current corpus of datasets as a baseline.

## Planned Workshops for 2013

As 2012 closes, the IMC is also planning workshops related to many aspects of the Network Information System (NIS) for the upcoming year. NIS production workshops planned for 2013 include:

- **Enhancement of the LTER Controlled Vocabulary to Support Data Synthesis**

The first version of the LTER Controlled Vocabulary was established in 2011, and term-searches in the Network data catalog take advantage of its synonym definitions. Several enhancements are already planned, particularly to include terms to more fully describe LTER data, and additional definitions and relationships for all terms. The group also will identify linkages to potential gazetteers (for place names) and to taxonomic authorities, and will plan software needs and implementation. They will also consider quantitative evaluation of term-use in datasets (e.g., EML “keywords”), and complete work on the draft Best Practices for LTER Keyword use in datasets.

- **GeoNIS Implementation with PASTA**

During 2013, the GeoNIS group plans a workshop to coordinate LTER Network Office (LNO) PASTA development and GeoNIS server administration, including guidance for a GeoNIS programmer and specifications for user applications and delivery of spatial data through the Network data catalog. Additionally, a group of site representatives will define workflows for ingestion of PASTA-contributed data into the GIS server, including conversion of base data to standardized layers, and creation of web mapping services.

- **Leveraging PASTA and EML-based Workflow Tools for LTER Data Synthesis**

Several 2013 cross-site synthesis projects are planned that will require data generated from EML-described data (e.g., ClimDB, Veg-DB, and Cross-Site Coastal Water Quality). Information managers plan a NIS workshop to leverage the PASTA framework and build reusable EML-based workflow software for these derived data products. This work will also provide crucial feedback to data providers and NIS developers, including practical real-world experience to inform recommendations for EML metadata content, congruence, and site data package management. This group also anticipates follow-up activities with scientists, students, and Network committees (IMC, NISAC and the Executive Board).

- **Managing sensor networks and data: Best practices for LTER**

Sensors are increasingly used in LTER research and LTER information managers often handle large data streams in near real time. In 2013, this group will develop a Best Practices guide for managing sensor data and networks. They will examine various strategies and applications in use both within the LTER Network, and from the broader environmental sensor community. Their recommendations will include tracking of sensor life cycle events and QA/QC procedures.

## Site Collaborations

Sites have continued their collaborations, and made many improvements to their local systems. In 2012, these collaborations accelerated at a nearly unprecedented rate, thanks to the injection of supplement money from NSF. In 2013, the availability of additional ARRA funds will support targeted activities focused specifically to enhance data availability:

1. broad-scale use of the Matlab tools for data processing and description from GCE
2. completion of the Drupal Environmental Information Management System
3. further support for inter-site consultation on EML construction and interpretation of PASTA quality checks

## A Busy Year Ahead

These activities, along with scientific synthesis working groups which will include IMC members, will mean a busy upcoming year. We all look forward to sharing advances and successes at our next meeting, to be held at one of our field stations in late summer 2013.

## Sensor Network at the Andrews Forest

edit

Adam Kennedy (AND), Don Henshaw (AND)

Researchers at the H. J. Andrews Experimental Forest LTER site maintain over 30 Campbell dataloggers for collection of climatological and hydrological measurements. Currently, about half of these dataloggers use existing radio telemetry to stream data directly to a base station at the Andrews Headquarters on an hourly basis. With the retirement of long-time systems administrator Fred Bierlmaier, and the hiring of Adam Kennedy, this past year has brought changes to the way these streaming data are handled. We have moved away from customized procedures to more fully embrace capabilities within the Campbell LoggerNet (LN) software. We have upgraded our LN server workstation, added a remote LN database to host raw data, integrated LN Real-time Monitoring and Controls to handle real-time data visualizations, and began sending auto-generated alerts to personnel on site. This updated toolset will enable the site to 1) seamlessly integrate with the LTER database located at Oregon State University (OSU), 2) immediately begin plotting highly customizable graphs online in near real-time, and 3) auto-notify information managers and field technicians of sensors that are in need of attention.

Andrews information managers were energized by information exchanges on sensor data at both the NERC/SensorNIS workshop at Hubbard Brook in 2011 and the recent ASM 2012 SensorNIS workshop at Estes Park. Site personnel are recoding stored procedures and other programs to schedule daily transformations from the raw streaming data in the LN database into our final online table formats. This transformation currently includes QC range checking and flagging of values, but other checks are being discussed. We are also restructuring our output tables to include coded attributes to indicate changes in methodology or instrumentation in collection of each measurement, and to include a quality level flag. The structure changes will also include revisions to our attribute names for better standardization and clarity of our hydrometeorological records.

In addition to these noted improvements in the workstation environment, the first phase of the new wireless communication network that extends deep into the forest - providing over 60Mbps of bandwidth to researchers in remote locations - is complete. This network consists of an array of point-to-point and point-to-multipoint wireless (5.8 GHz, 2.4 GHz, and 900 MHz) links. This wireless backbone and associated cloud radios will support long-term research, high-resolution sampling campaigns, dynamic interpretative trails, and virtual classrooms.

## Sensors at North Temperate Lakes

edit

Corinna Gries (NTL), Aaron Stephenson (NTL), Ken Morrison (NTL)

In collaboration with the Global Lakes Ecological Observatory Network (GLEON), NTL is currently maintaining six lake buoys and one terrestrial weather station. Each buoy is equipped with a thermistor chain measuring water temperature at one meter intervals, and measures basic weather parameters above the water (air temperature, wind speed and direction, relative humidity, etc.). Other parameters measured on some buoys include photosynthetic active radiation, dissolved oxygen, conductivity, chlorophyll concentration, and phycocyanin concentration. One buoy is equipped with moving sensors that traverse the water column continuously. A single buoy can deliver up to 40 distinct data streams for the deepest lake; in total, ca. 140 data streams at one minute intervals are recorded. The sensors are controlled by Campbell CR1000 dataloggers which also function as temporary data repositories for up to four weeks of data. The network of dataloggers is configured and data is retrieved from these loggers using Loggernet, Campbell Scientific proprietary software. Communication between the point-of-presence computer running Loggernet and the field-deployed Campbell dataloggers uses a 900MHz radio network based on Freewave full-spectrum radios. The data are pulled every hour from the data loggers onto the point-of-presence servers. For one season, (May through November) one buoy can collect up to 60 MB worth of data.

A DataTurbine server is running to allow for easy sensor health monitoring. The data are not currently streamed into the database directly by DataTurbine; instead the data logger files are manually uploaded into temporary tables. Database triggers then apply range checks before data are archived, with appropriate QA/QC flagging, in their respective final tables. A web application that accesses these final tables allows users to query, subset, and download the specific data they want.

Several aspects of this approach, plus various custom applications (e.g. Ziggy data loader), have been in operation at NTL for many years. We are now in the process of updating the system with NSF and Moore Foundation funding to Tony Fountain at CallT2. Additionally, the GCE Matlab toolbox will be incorporated into our system to work seamlessly with DataTurbine for quality control of streaming data. Also, another buoy has recently been installed which communicates data between two DataTurbine servers, one on-site and one at the point-of-presence server, eliminating the need for a datalogger. It can be expected from these projects that by the end of 2013 the installation and maintenance of DataTurbine will become more user friendly. Currently unresolved issues of concern are related to tracking the type and make of sensors, calibrations, and other maintenance activities.

## Sensor Activities on the Virginia Coast Reserve

edit

John Porter (VCR)

Since the SensorNIS workshop, sensor activities at the VCR/LTER have revolved around several different activities.

First, the deployment of some new sensing systems including a radar-based tide gauge was deployed on Hog Island. This gauge sends data wirelessly using a Campbell Scientific CR206, our main "workhorse" logger these days, to a network node on the north end of Hog Island and then sent via Wi-Fi back to our lab. Additionally we have been deploying networks of autonomous sensors that are too isolated to reach our existing network backbone. These include ground water monitoring stations on Smith and Metompkin Islands, and a network of tipping-bucket rain gauges deployed along the Delmarva Peninsula. Although all these stations could be accessed wirelessly, the cost of doing so would be high; so at least for now they will be dumped manually. Most of these stations are deployed where they can be reached fairly easily by car or after a short boat ride.

Second, we have been taking a harder look at how to handle reporting of sensor problems and the creation of level-1 datasets that have more advanced QA/QC and data flagging. We are moving away from a system that had two fundamental data forms to one that uses three fundamental forms. The "two-form" system had raw data in the form it came in directly from the sensor, typically as a text file. This was then processed to create a rudimentary level-1a dataset by ingesting the data and doing basic data type and range checks. Any additional corrections, such as correcting clock errors or sensor calibration errors, were made by altering the level-1a product using a program. The downside of this model is that post-hoc corrections need to be very carefully applied because once a number has been corrected (e.g., multiply by 2), you don't want to accidentally run the same correction again (e.g., multiply by 2 again, yielding a multiplication by 4 of the original data). In the "three-form" model no post-hoc corrections are applied to the low-level level-1a data table. Instead, a program reads the level-1a data and applies post-hoc corrections to produce a level-1b dataset. The level-1b dataset is repeatedly recreated by the program, re-applying the needed corrections. It is possible to eliminate the intermediate (level-1a) data if you go directly from the raw data to the corrected data using a program that re-ingests and re-applies corrections. However, many of our datasets have gone through a progression of raw forms, many of them incompatible, so that there is a significant advantage to processing them after ingestion.

We have also been working on a database for reporting sensor problems. The draft web forms allow users to select a type of station, identify a particular station and the sensors affected, and to recommend actions to be taken. The database will then be used to automatically write corrective code to flag or remove problem data.

### Commentary

---

## Targeted searches with EML and LTER Controlled Vocabulary

edit

Margaret O'Brien (SBC)

This essay will illustrate one way in which these two structures - EML datasets and our SKOS vocabulary - can be used together right now to further improve a user's experience when looking for LTER data. The examples here could be implemented in our current catalog; they do not require PASTA. Definitions for terms can be found at the end of this essay.

### Background

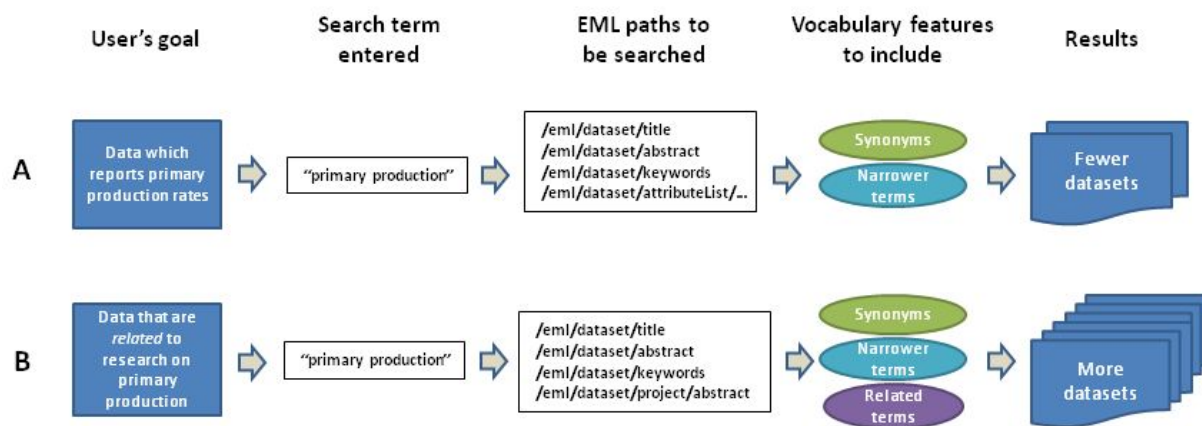
One of the reasons we adopted EML as the format for LTER metadata was that it was structured. The metadata structure (XML path) where a term occurs carries information in addition to the term itself, so searches can take advantage of that structure by accessing specific metadata components such as dataset/abstract, or dataset/creator. The LTER data catalog's queries have used EML's paths for many years. In 2011, the LTER Controlled Vocabulary project began structuring

our search terms in a format called SKOS, arranging terms into hierarchies with relationships such as synonymy, and the network catalog now uses the SKOS structure to drive an auto-filled form for searches by term. But the current network catalog does not yet take full advantage of these structures.

## The Core Areas Search Challenge

Our audience needs the ability to “find data for our core research areas”. But our research is interdisciplinary by nature, so a single dataset is often related to several research areas. To further complicate matters, some terms, like ‘primary production’, are both a measurement (i.e., areal uptake of carbon over time) and a topic of study. So there could be two ways to interpret the request “show me data for primary production”. As illustrated in the figure below, the user might want either A) data reporting production rates, or alternatively, B) data *related* to research on primary production. Obviously, two different queries should be offered, and the concept ‘*related*’ is crucial to one of them. Fortunately, with structured EML and a structured vocabulary, it is already possible to build these.

Figure 1. Example of two searches, where each is targeted at a specific type of data request.



The two query types take advantage of different features of the SKOS vocabulary, and search different parts of the EML. By designing distinct queries that are clearly labeled and have appropriate search parameters, the possible uses of the same term (‘primary production’) can be clearer to the user. A system such as this separates the catalog’s responsibility from the data’s. The data package does not need to ‘know’ what research projects might use it, but the system does. The EML content is the responsibility of individual sites, scientists, and information managers, while the Vocabulary (as part of the ‘catalog’) is the responsibility of the Network.

## Requirements

To achieve the desired results, we need two things:

1. A vocabulary that makes all the proper linkages and contains the expected terms to be used for all LTER data. It will be particularly important to make connections between related terms.
2. Data that are described explicitly and carefully. The EML path ‘dataset/abstract’ must describe only the data, and other details about the scientific project that generated it are in their appropriate locations, for example, ‘dataset/project/abstract’. Keywords should apply to data only, and not to the projects that use them. For example, if a dataset is of carbon dioxide measurements it should *not* have the keyword ‘primary production’. The linkage between ‘carbon dioxide’ and ‘primary production’ is taken care of by the Vocabulary.

## Limitations

What these searches cannot do:

1. They do not group together datasets that are related to a specific site-based or network-based research project. That could be accomplished with queries, but different from the examples above.
2. They cannot make inferences about the appropriateness of data or a particular use. For that functionality, we need more sophisticated knowledge models such as ontology.
3. These example queries are still based on simple string-matches in the EML. So any dataset that uses the term ‘primary production’ in a searched field will be returned (e.g., the phrase “data describe the transect for our primary production study”), and would be false positive for query type A. To reduce those false positives, we would need a more complex annotation system between the EML and the catalog.

## Conclusion

Designing a few targeted queries is not a major or sophisticated change to the current catalog. It can be accomplished with the Controlled Vocabulary as it stands now, and can be applied to either a Metacat back end, or the developing PASTA API. Currently, we have only one term-based search form in the catalog. It appears to be of query type A, and it’s generally parameterized that way. However it returns results that are closer to the expected results of query type B. This may be due in large part to inappropriate keyword use in datasets. As with many uses of EML, complete analysis may indicate that EML paths other than those listed in Figure 1 should be considered.

## Definitions

**EML Path:** the location of a metadata-item in the EML document, e.g., ‘/eml/dataset/title’ is the XPath to the data package’s title.

**LTER Controlled Vocabulary:** a set of terms structured into SKOS. The vocabulary can be browsed here: <http://vocab.lternet.edu> (Porter, 2010, 2011)

**Synonym:** a term in the LTER controlled vocabulary that can be used in place of another term. For example, ‘nitrate’ and ‘NO3’ are synonyms.

**Related terms:** terms in the LTER controlled vocabulary that are not synonyms, but that could be included to expand a search. Some groups of related terms include 'carbon, NPP, primary production', or 'nutrient flux, nitrate'.

## References

Porter, J. 2011, Managing Controlled Vocabularies with "TemaTres". Databits, Spring 2011, <http://databits.lternet.edu/spring-2010/controlled-vocabulary-lter-datasets>

Porter, J. 2010., A Controlled Vocabulary for LTER Datasets. Databits, Spring 2010, <http://databits.lternet.edu/spring-2010/controlled-vocabulary-lter-datasets>

## A Request for Mr. S. Claus - Some Sensor Tools!

edit

THE FOLLOWING LETTER WAS FOUND IN A SNOW DRIFT OUTSIDE THE NSF HQ EARLY LAST WINTER NEAR WHAT APPEARS TO BE A REINDEER HOOF TRACK.

Dear Santa,

I know that sensors pose a challenge for you in your work (especially "roof-cams"). For Christmas, some people might ask for new sensors, but I'd like some software tools for managing them. In particular, I'd like:

1. **Improved capabilities for tracking sensor events**, including deployment, calibration and failure. When I started with the basic sensor network you bought me 3 Christmases ago, it wasn't hard. First, there weren't so many different sensors. Also, they were all new, so they worked great. However, now with the advanced sensors you've brought me on subsequent Christmases, I'm getting overloaded! I'd like the data about the sensors and their deployments to be in a database so that I could:
  1. Use queries to identify sensors that may be reaching the end of their useful life so that I can replace them before they fail.
  2. Use queries to identify time periods where sensors were inactive or providing less-than premium data so that I can put the appropriate "flags" in my datasets.
  3. Provide a basis for statistical analyses focused on identifying causes of sensor failure, so that I can anticipate future problems.

Please make it so that populating the database doesn't require too much labor in and of itself. When I or my technician is in the field, it's not always the best time to record copious notes, especially when we are rushing home to wrap gifts or bake cookies.

2. **Better tool sets for identifying sensor problems**. As you know - especially following the unfortunate episode of the "defective Rooty-Toot-Toots" - the key to avoiding big problems is early detection of small ones. I can generate some basic QC tools, but it would be great if you could integrate a bunch of standard tools into a set where I could easily apply them to new datasets. John C. and Don H. have been talking about developing some standards, and it would be great if you could integrate those into some tools that I could easily use with minor configuration for new datasets.
3. **More sophisticated image and soundscape analysis tools**. As a world traveler of some renown, you know how valuable sight and sound are. After all, if it weren't for your "sees you when you're sleeping" device, you'd have been "busted" years ago for burglary. However, unlike you, we don't have an unlimited supply of elves to watch and listen. It would be great to have some programs that would automatically extract information from image and sound streams.

I know that a lot of other information managers will want these kinds of things, too, because they all have sensors, sometimes the same kind, sometimes different, some old, some new. So Santa - this line of toys will make lots of people happy. I have been a good information manager all year, and have faithfully worked on improving my metadata, so I'm hoping you and your elves will come through with the goods!

Sincerely,

NAME TOO SMUDGED TO READ

## News Bits

---

## Sensors and Superstorm Sandy

edit

John Porter (VCR)

As this issue of Databits was about to go to press, "superstorm" Sandy had a near-miss on the Virginia Coast Reserve. It provided an excellent opportunity to see how sensors perform under heavy-duty conditions. Generally speaking our sensors worked well throughout the storm, and it appears that all of them on Hog Island (our principal barrier island research site) survived and collected interesting data throughout the storm.

This is not to say that Murphy's Law didn't come into play as well. Our main radio hub on the north end of Hog Island developed a power glitch, probably related to a dodgy power inverter, two days prior to the arrival of Sandy that caused it to come up for about 1 minute, transfer lots of data, then go off the air again for the day. This was annoying in that it prevented us from dumping the data from the island real-time, but even the brief transfer window was enough to get back some interesting results. Also in accordance with Murphy's Law, now that the storm has past, the radio hub is working perfectly!

Our biggest suspense was over a new radar tide station installed on an abandoned navigational marker only days prior to Sandy's arrival. The box containing the sensor and logger are suspended about 3 meters above the water, but we feared that a major storm like Sandy could produce a large storm surge that might just be high enough to wash the box away! Fortunately, as seen in the graph below, the water level only rose to within 1.2 meters of the box so we got a good first-

hand view of the storm surge related to Sandy. Especially notable is how quickly the storm surge abated as Sandy passed. Tides returned to normal ranges over the course of a single tidal cycle, dropping over 2 meters between high and low tide (normal tidal range is about 1.3 m), after building up for several days as Sandy approached.

Another source of suspense was our network of ground water wells. These wells are located in rough lines across Hog Island, and although all the electronics are located well above ground, the exceptional tidal flooding still placed them at risk. Fortunately, none of the data loggers appear to have been submerged, although one located by a pond probably came pretty close as water levels "spiked" around the time of peak flooding from Sandy. Note on graph below, the spike represents standing water around the well - and the quick decline occurred as water flowed away above-ground. The slower decline on the "shoulder" of the curve represent the slower flows out of the well through the sand.

Sandy also provided an excellent opportunity to test out some of the newer sensors we have deployed. One is an "impact" rain gauge that uses the impact force of rain drops onto a piezoelectric sensor to model the rain amounts. We are hoping to replace many of our (frequently clogged) tipping bucket rain gauges with this technology. The *Morella cerifera* seeds that abound on our site are exactly the right size to get through the mesh on a tipping bucket rain gauge, but just big enough to clog up the drain! However, the flat plate of the impact gauge can't be clogged. To test the impact gauge, our newest station on Hog Island includes both tipping bucket and impact gauges. During Hurricane Irene, the bucket blew out of the tipping bucket gauge, so we weren't able to do a comparison. However, during Sandy both gauges stayed operational and the correspondence between sensors was excellent. In the graph below, the red line indicates the tipping bucket rain gauge and the blue the impact gauge.

Finally, no discussion of sensors at the Virginia Coast Reserve LTER would be complete without mentioning the webcams. Again, Murphy's Law kicked in with our Broadwater Camera experiencing a power problem (again, probably a bad inverter) that caused it to communicate much more slowly than usual, with "ping" times of over 1 second. Nonetheless the camera was able to capture an array of photos during the storm. Many of them were less than optimal because the high winds caused the camera to shake, but they still provide a graphic record of the extent of flooding during "superstorm" Sandy. Below are pictures of South Hog captured during peak flooding and after flooding subsided. The dark areas are the *Morella* shrubs that stuck up out of the water. Also there is a picture of a very wet Peregrine Falcon that rode out the storm, with its mate, in the Cobb Island hacking tower. Additional photos are available at: <http://amazon.evsc.virginia.edu/gallery23/main.php/v/events/Sandy2012/>

## SensorNIS: Community engagement to build a resource guide for managing sensor networks and data

edit

Don Henshaw (AND), Corinna Gries (NTL), Renee Brown (SEV), Jason Downing (BNZ)

LTER Information managers and site researchers are actively developing sensor networks and supporting information systems. Sensors are increasingly used in LTER research and information managers are expected to manage large, near real-time data streams. As sites develop system software and management protocols to accommodate these sensors there is need for coordinated efforts to help build agreement on general strategies and provide training for handling these large data streams. The willingness of site personnel to share expertise and explore possible solutions has been evident in several activities over the past year. Strong interest in information exchange workshops and training sessions both within the LTER Network and throughout the broader environmental community demonstrate the pressing need for common strategies and shared resources. This article reviews some of these activities including the recent workshop at the LTER All-Scientist Meeting (ASM), and invites your participation in these ongoing efforts.

The need for information exchange regarding sensor network expansion within LTER was targeted by NSF supplemental grants to LTER sites in 2011 and SensorNIS was born. SensorNIS funding to sites in co-sponsorship with the Northeastern Ecosystem Research Cooperative (NERC) supported the "Environmental Sensor Network / LTER SensorNIS Workshop" at Hubbard Brook Experimental Forest LTER, New Hampshire, in October 2011 (<http://databits.lternet.edu/fall-2011>). A pre-workshop survey indicated that existing site sensor systems did not meet site needs in terms of acquiring, handling, providing quality control, and documenting high volumes of incoming streaming data. The workshop focused on management of sensor networks and quality control (QA/QC) of incoming data streams (<http://im.lternet.edu/projects/SensorNIS>). Thirty LTER representatives were among the seventy-two participants. Recommendations from this workshop included increasing training opportunities and developing a web-based resource guide or best practices covering many aspects of sensor network establishment and management through community participation.

Subsequently, the LTER Network Office hosted the 2012 cost-shared training workshop, "Software Tools for Sensor Networks" (<http://news.lternet.edu/Article2590.html>), sponsored by the LNO, NCEAS, and DataONE. This training workshop demonstrated multiple software tools in the handling and managing of sensor data including acquisition, transport, raw data storage, QA/QC, and archival (<http://im.lternet.edu/node/999>). Highlighted software tools included GCE Matlab Toolbox, Kepler, DataTurbine, the CUAHSI software stack, and R. Additionally, the Sevilleta LTER provided the "Data Acquisition from Remote Locations" training workshop to demonstrate existing experiments with telemetry operations and to provide hands-on training in basic electronics, photovoltaic systems, Wi-Fi telemetry, data loggers and basic programming. The number of applicants for both workshops far exceeded the training room capacity and similar training sessions have been proposed for 2013. Moreover, results from both of these training sessions indicate the need for better information exchange and common resources throughout the broader community.

The ASM 2012 workshop, "SensorNIS: Building a sensor network resource guide through community participation", was intended to build on the results of the NERC/SensorNIS Workshop and subsequent training sessions in developing a sensor network resource guide. The initial outline for such a resource guide was presented by workshop organizers and has been modified based on the review and feedback from the twenty-five participants ([http://im.lternet.edu/resources/im\\_practices/sensor\\_data](http://im.lternet.edu/resources/im_practices/sensor_data)). The intent of the organizers is to engage the environmental sensor community to identify potential sources of information, solicit and assemble contributions, and enlist editors to moderate each topical section of the resource guide. The planned topical sections are:

- Sensor, site and platform selection
- Data acquisition and transmission
- Sensor management, tracking, and documentation
- Streaming data management middleware
- Sensor data quality assurance and quality control (QA/QC)
- Sensor data archiving

Suggestions for how to approach building the resource guide were made by workshop participants and fulfilling these recommendations will rely primarily on community engagement. A proposal for a product-oriented working group was submitted, titled "Managing sensor networks and data: Best practices for LTER", and is currently under review. Other information gathering efforts geared toward building the resource guide include this issue of DataBits, which is collecting narratives of different sensor implementations at sites, and specific site surveys, which are planned to learn about sensors, platforms, data acquisition and transmission approaches.

In addition to these more LTER-centric activities a user group is currently being formed under the umbrella of ESIP (Federation of Earth Science Information Partners <http://esipfed.org/>). ESIP provides basic support for such a user group, that is, it provides a managed e-mail discussion list, web space for the resource guide, teleconferences, and an annual meeting where members of this user group may get together. ESIP was chosen because many participants of the aforementioned workshops and trainings came from a wider community of environmental sensor users and expressed their interest in participating in this activity. An invitation to this new user group will be forthcoming and we welcome your participation!

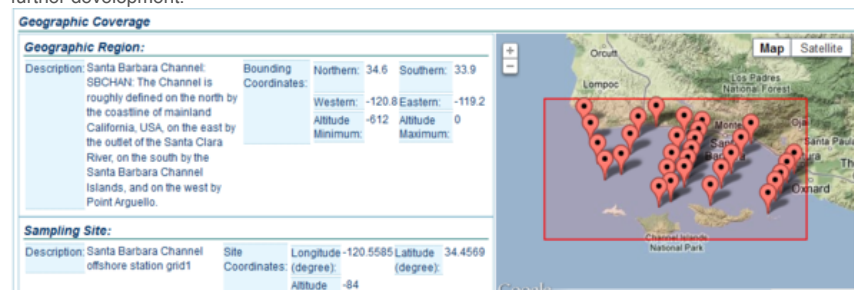
## EML and Google Maps

edit

Margaret O'Brien (SBC)

The Santa Barbara Coastal LTER's data catalog now includes a Google map generated from EML dataset-level <geographicCoverage> elements. The map required only one additional XSL template that mixes XSL with JavaScript; bypassing the need for KML files. XSL statements cycle through the <geographicCoverage> elements and define JavaScript variables, and all <geographicCoverage> nodes under <coverage> are plotted on the same map. All the JavaScript is based on examples available with the Google Maps API (v3). This code has been implemented at SBC and MCR, and could be easily incorporated into another HTML display of EML metadata. The below map shows 25 sampling sites and a bounding box derived from metadata from one SBC dataset (knb-lter-sbc.1016).

Currently, the code uses the EML-required elements under <boundingCoordinates>, and logic determines whether a location should be displayed as a single point (marker) or a box (polygon). An optional <gPolygon> can also be plotted (not shown here). Some additional features are straightforward to add, such as dynamically setting the map's center and zoom. Accommodating other <geographicCoverage> elements at other locations in the EML document will require further development.



## Good Tools And Programs

## A Web Service for EML-based Mapping

edit

John Porter (VCR), David Richardson (VCR)

The VCR/LTER has published a new web service that automatically creates a Google Earth KML file containing markers for each of the locations identified in an Ecological Metadata Language (EML) document. The tool returns a KML file named "EMLdoc.kml" that is suitable for use with Google Earth, ArcGIS or any other software that takes KML files as input.

The draft web service is being used to "feed" the coverage display on the VCR/LTER data display (<http://www.vcr.lter.virginia.edu/cgi-bin/showDataset.cgi?docid=knb-lter-vcr.25&displaymodule=coverageall>).

The display includes an openLayers JavaScript viewer that displays the points from the KML document generated by the web service. These sorts of displays can be easily exported to other LTER sites by making only a few changes to the JavaScript code.

David Richardson is continuing work on improving the web service to make it display more clearly the geographic entities that are described by bounding boxes or polygons (currently only centroids are displayed). There are also plans to move the web service to a more REST-type interface. However, it has been released for use by early adopters because it is functional and stable now.

The draft service for a package in the LTER or KNB Metacats:

[http://www.vcr.lter.virginia.edu/data/eml2/getKMLfromEML.php?knb\\_package=knb-lter-vcr.25](http://www.vcr.lter.virginia.edu/data/eml2/getKMLfromEML.php?knb_package=knb-lter-vcr.25)

The service for an arbitrary EML document can be specified by URL:

<http://www.vcr.lter.virginia.edu/data/eml2/getKMLfromEML.php?emlURL=http://www.vcr.lter.virginia.edu/data/query/text/eml/VCR97018.xml>



## Good Reads

---

### IT Support and GIS/Remote Sensing Analyses are Vital for Disaster Rapid Response at UNOSAT

edit

Using GIS and Remote Sensing software and data, and relying on 24/7 IT support, analysts at the United Nations Operational Satellite Applications Programme (UNOSAT) provide vital and timely information to disaster rapid response and relief organizations. This article details work that UNOSAT does, how they get notified of a disaster, how they get their source data, how they produce useable information for other organizations, and how important their IT support is to their work.

<http://www.isgtw.org/feature/satellites-servers-and-story-telling-help-disaster-rapid-response>

### Data Reporting and Data Usage, European Style

edit

Riebesell U., Fabry V. J., Hansson L. & Gattuso J.-P. (Eds.), 2010. Guide to best practices for ocean acidification research and data reporting, 260 p. Luxembourg: Publications Office of the European Union.

Although this guide was written specifically for Ocean Acidification data, Part 4: 'Data reporting and data usage' applies equally well to ecological data in general.

Each section begins with a quote summarizing a data management challenge. All will sound familiar to an LTER site IM, such as "Organising and documenting data in order to meet the requirements of data archives requires time and efforts, as with any other media used by scientists to communicate their findings, e.g. scientific papers, posters or oral presentations." Then the section proceeds to address that challenge, such as "First, research programmes must allocate funding to data management. Each program should hire a person to create metadata, contact scientists to prepare and submit their data, aggregate datasets that are related but come from different sources, and submit/import data into a database."

Part 4 covers the following topics: sharing data, safeguarding data, harmonizing metadata with data, reporting and disseminating metadata and data, pitfalls to avoid and recommendations for standards and guidelines. Much of the advice will sound familiar to an experienced IM but is worth reading if only to recognize how a completely separate data management group in Europe describes and advises on the same challenges as we face in the LTER.

Summary Document here:

<http://www.epoca-project.eu/index.php/guide-to-best-practices-for-ocean-acidification-research-and-data-reporting.html>

This book is available as a PDF at this link:

[http://epoca-project.eu/index.php/restricted-area/documents/doc\\_download/658-guide-to-best-practices-for-ocean-acidification-research-and-data-reporting.html](http://epoca-project.eu/index.php/restricted-area/documents/doc_download/658-guide-to-best-practices-for-ocean-acidification-research-and-data-reporting.html)

ISBN: 978-92-79-20650-4

DOI: 10.2777/66906

