Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-03-22 17:13:16
REFERENCE NO: 190

Author Names & Affiliations

- Joel Widder - National Association of Marine Labs

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

ocean, coastal, and Great Lakes research and education

Title of Submission

Cyberinfrastructure Needs of Marine Laboratories

Abstract (maximum ~200 words).

This submission provides relevant excerpts from a report from the National Academies entitled: Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the 21st Century. As part of this overview of challenges facing field stations and marine laboratories, this report contains a discussion and recommendations regarding the cyber infrastructure needs of marine laboratories and field stations.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

The rapid environmental changes that are taking place globally raise basic research questions and present major societal challenges. Evidence is mounting that the growing human footprint is stressing natural and social systems. Climate change, biodiversity loss, natural resource extraction, and pollution pose considerable threats to ecosystems, economies, and human well-being. Coping with the challenges will require improved knowledge about the social–ecological system. Field observations have played and will continue to play critical roles in the physical, natural, and social sciences.

Field stations are national assets formed by the unique merger of natural capital, intellectual capital, social fabric, and infrastructure that leads to the important scientific endeavors required if we are to understand our rapidly changing natural world. Field stations, either inadvertently or by design, are repositories of long-term observations and datasets of natural history necessary for documenting global changes. A greater emphasis on integrated, multidisciplinary research that includes the physical sciences, geosciences, social sciences, humanities, and arts will enhance scientists' use of historical datasets to address global challenges. The recognition of the importance of this portfolio of activities in what is now called "convergence"3 is a strength of many field stations.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Challenges of Maintaining and Upgrading Infrastructure
Field station managers and users have long recognized the need for safe, functional housing and properly equipped workspaces. Their two primary challenges in this regard are maintaining aging facilities and keeping up with rapid advances in technology. The latter is particularly important because laboratory- based research is increasingly integrated with field research. All infrastructure requires preventive maintenance, replacement, upgrading, or some combination of the three. That is not peculiar to field stations. However, the sites in which field stations are embedded and that make them attractive—along coasts, in mountains, in forests, or in deserts—often expose their infrastructure to extreme, highly variable environmental conditions that can take a toll. In addition, many stations are located in areas that are vulnerable to wildfires, earthquakes, tornados, hurricanes, or other natural hazards. These vulnerabilities add to the cost of maintaining the facilities and pose a risk to research equipment (e.g., laboratory equipment such as microscopes, autoclaves, and ultra-cold freezers, and field equipment such as nets, boats, and environmental sensors), biological collections, and data stored at these stations. Field station facilities can degrade much more rapidly than equipment found in environmentally controlled laboratories, and this is a financial burden on field station managers and in some cases compromises the research.

The recent survey by the National Association of Marine Laboratories (NAML) and the Organization of Biological Field Stations (OBFS) reveals common infrastructure priorities among field stations (NAML-OBFS 2013b). The top priorities are electricity, Internet access, support staff, laboratory space, storage, long-term monitoring, classroom capacity, housing, on-site maintenance, and engineering capacity. Respondents suggested that increased support for Internet access would improve scientists' ability to use field stations while providing potential visiting scientists with access to specific data catalogs that are critical for developing research programs. According to the survey, a major problem is that basic data catalogs—species lists, maps, weather data, and land-use history—often are lacking at field stations. Some respondents indicated that field stations had insufficient space for laboratories, classrooms, and storage (including refrigeration). Data-management systems were considered excellent by a few respondents but ineffective by others. In addition, field researchers may require transport to and from field sites. Transport needs can vary from a golf cart to a submersible, depending on the site and the research being conducted. Field stations with increasingly sophisticated scientific equipment and automated sensors will also have to make investments in the capacity to capture, process, store, and share increasingly large datasets. Consideration must also be given to data that do not typically lend themselves to classic deposition in accessible databases, such as video recordings of animal behavior or deep-sea observations. Upgrading data- management systems was also identified as a high priority in the survey.
Investments in maintaining existing infrastructure clearly are a primary concern for field station administrators but often are a relatively low priority for their host institutions, particularly if a field station is remote. Only 14 percent of respondents (N = 197) to the NAML-OBFS survey noted that financial planning for field stations included depreciation of buildings and equipment. That result is a remarkably low percentage, considering the respondents' overwhelming sense of vulnerability to anticipated funding losses in operational revenue (76%) and in federal (65%), state (60 %), administrative (54 %), and donor (54%) support over the next 5 years.
There is clear need for every field station to develop a comprehensive infrastructure-management plan that is integrated with its strategic mission, its science plan, and its business plan (Lohr 2001).

Cyberinfrastructure and Connectivity
The inclusion of data as a type of infrastructure represents a paradigm shift for many field stations. Data constitute a primary product of field stations; if these data were made easily available, they could serve a broad audience. Long-term and baseline natural-history data should be an attraction to scientists and educators and be counted as part of a field station's value (see Chapter 6), and move them from serving merely as environmental sentinels to active participants in solving ecological and economic problems at a variety of scales. The acquisition of data is only one part of the equation. Data must be stored, managed, and integrated to ensure that they can be mined, visualized, and accessed through high-performance Internet connections—all parts of the domain of cyberinfrastructure.
Cyberinfrastructure consists of the assortment of information technologies that enable data storage, management, integration, and analysis. It is increasingly recognized as essential to science in that it dramatically improves scholarship and research productivity. Efficient cyberinfrastructure generally requires reliable Internet connectivity and modern computer hardware. At a minimum, field stations need adequate Internet connectivity to facilitate user access and collaboration. The availability of adequate cyberinfrastructure attracts scholars who are interested in cross-disciplinary research and fosters new scientific endeavors in emerging fields. Every field station should provide—whether on site, at a selected hub location (such as a host institution, the National Ecological Observatory Network (NEON), or other research centers), or through a collaborative network—online access to the complete historical datasets of its natural and human history and provide means by which its users can contribute to these datasets. This type of interactive access to databases can provide quality control of data in that scientists can monitor data input and output in real time and respond to anomalies. Scientists, students, or

even visitors can see how data that they collect fit into larger temporal and spatial contexts.

Data Management
Infrastructure to organize, archive, and share data collected at a field station could expand the impact of a field station's research by making data available to other researchers to use, and by facilitating the ability to track data use and impact. Many tools for ecological data storage and recovery have been developed by the Long Term Ecological Reserve Network (LTER), the National Center for Ecological Analysis and Synthesis (NCEAS), the Knowledge Network for Biocomplexity (KNB), and others. Ecological metadata language developed by KNB and NCEAS has been widely used and is compatible with the larger aggregators (such as DataONE and Google). The National Park Service (NPS) has a research permitting and reporting database and a website that allows investigators to request reports and research data from specific national parks. The NPS website can be searched by park, taxon, or investigator. Sharing the data products from field stations broadly would add value to the data and to the field station where the data were collected. Without centralized repositories, data developed at field stations are easily lost. Alternatively, if they are archived and made widely available, they have ever-increasing value to provide perspectives on environmental change. Archiving and sharing data from field stations are critical.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

The Dark-Data Problem
Researchers at field stations often record data in their logbooks and spreadsheets either by hand or electronically. They take the data with them and, historically, rarely share them with the field stations where they conducted their research. Some of the raw data are eventually analyzed and the results incorporated into peer-reviewed publications; some may be lost when a researcher is no longer active, and these fall into the realm of "dark data34"—data that are inaccessible to the broader scientific community that relies on new, more sophisticated data-management tools. Salvaging the large body of historical dark data that still reside in notebooks, file cabinets, or memories of aging investigators is a challenge, but worth pursuing. The Berkeley Ecoinformatics Engine,35 funded by the Keck Foundation at $3.5 million, could serve as a model for addressing both the dark-data challenge and the problem of integrating diverse databases within regional networks of field stations. The intent of the program was to organize and unify the wealth of data in University of California, Berkeley laboratories, natural-history museums, and field stations and to merge them with diverse environmental baselayers on climate, land cover or use, vegetation indexes, hydrology, and fire and other freely available datasets. The results are available for rapid exploratory analyses, tests for correlation, and visualizations that communicate results to a broad community of users. The Ecoinformatics Engine unites previously disconnected perspectives from Earth and atmospheric scientists, geographers, paleoecologists, and ecologists and enables tests of predictive models of global change. This constitutes a critical advance in making the science more rigorous.

Scaling Up and Sharing -- Field station cyberinfrastructure is physically and technically diverse—from digital sensor arrays to high-speed communication networks—and varies widely among field stations. Because of the diversity, a comprehensive infrastructure-management plan is best constructed around broad categories of use rather than type (physical, technical, and cyberinfrastructural). Three such use categories, modified from those described by ESFRI for large-scale research infrastructures (such as that of CERN, the European Organization for Nuclear Research), are (1) single sites, including infrastructure on the site of the field station itself; (2) networked sites, distributed resources and databases and infrastructure that are shared, possibly through collaborative networks; and (3) global infrastructure, available through online networks. Those categories could be used to outline infrastructure in the context of individual field station needs and services. For example, the infrastructure- management plan would describe how and when data collected with a place-based infrastructure are to be stored (remain part of the field station infrastructure) and how and when they are to be shared in a distributed framework (a service provided by the field station to the scientific community).

Sharing information and resources among field stations is critical in a world in which technological advances and expenditures increase at a rapid rate. The resources may include datasets on soil types, land-use history, climate, and aspects of biology. Through networking, field stations and researchers can share resources and collaborate on common topics and scientific questions. Sharing of data requires use of standard formats and metadata.

When current best practices for data storage and metadata registry at the network level are used at a field station, they can become a part of a much larger, global infrastructure as modern data aggregators and information-management tools develop (e.g., DataONE and Google). As future information technology allows greater access to multiple data sources, the need increases for uniform data collection on target organisms and environmental properties and processes to allow analyses on regional and national scales.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).”

Links to reports mentioned below:

- NAS - Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the 21st Century (2014), https://www.nap.edu/catalog/18806/enhancing-the-value-and-sustainability-of-field-stations-and-marine-laboratories-in-the-21st-century

- Field Stations and Marine Laboratories of the Future: A Strategic Vision, http://www.obfs.org/assets/docs/fsml_final_report.pdf.

From: Joel Widder
Sent: Wednesday, March 22, 2017 11:40 AM
To: NSF CI 2030 RFI <nsfci2030rfi@nsf.gov>
Subject: Re: NSF Advanced Cyberinfrastructure, Request for Information

On behalf of the National Association of Marine Laboratories (NAML) and in partial response to the request for info on cyber infrastructure, we would like to bring to your attention two relatively recent reports that examined the future infrastructure (and other) needs of marine laboratories and field stations. We have attached copies of these two reports: one is a workshop report supported by NSF/BIO; and the other is a National Academy of Sciences report on the future of marine labs and field stations (also supported by NSF/BIO). These reports both include a discussion of cyber infrastructure challenges and opportunities facing the research and training activities of marine laboratories. We hope they will be a useful resource in response to the RFI.

Joel Widder
Co-Founder and Partner
Federal Science Partners
1201 New York Avenue NW
Suite 400
Washington, D.C. 20005
www.FederalSciencePartners.com