Characterizing the Utilization of Large Scientific Research Facilities:

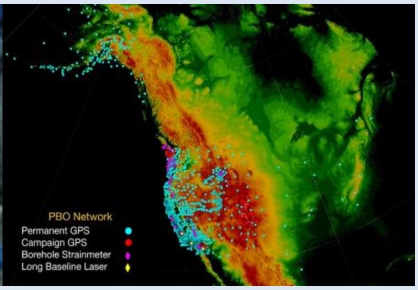
An Analysis of Users and the Evolution of Use at NSF-Supported Multi-User Facilities

Kristin Ludwig, Ph.D.

AAAS Science & Technology Policy Fellow
National Science Foundation

August, 2012



Characterizing the Utilization of Large Scientific Research Facilities:

An Analysis of Users and Evolution of Use at NSF-Supported Multi-User Facilities

Kristin A. Ludwig, AAAS Science & Technology Policy Fellow
National Science Foundation – Large Facilities Office

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author greatly appreciates the generous support of both NSF and AAAS throughout this fellowship. She thanks Bill Miller for his consistent guidance and encouragement and her Large Facilities Office colleagues – Mark Coles, Chuck Bouldin, Patti McNamara, Florence Rabanal, and Phil Schwartz – for insightful conversations and constructive feedback during this study. She thanks Greg Anderson, Alex Isern, Joy Pauschke, Jeff Pier, Robert Robinson, and Jean-René Roy for useful comments that improved this manuscript and gratefully acknowledges the valuable input from colleagues at NSF, STPI, OSTP, STScI, and DOE on this project. Finally, the author thanks the interviewees at the case study facilities and awardee institutions for their generous time and interest: this work would not have been possible without their input.

AUTHORSHIP AND STUDY PROVENANCE

This work was completed as an independent study solely by its author, a science policy fellow in residence at the National Science Foundation's (NSF) Large Facilities Office (LFO) from September, 2011 through August, 2012. The author was supported by an American Association for the Advancement of Science (AAAS) Science and Technology Policy Fellowship for the duration of this study. As an AAAS Fellow, the author does not represent either the NSF or the AAAS. The contents of this report do not represent official positions of NSF, AAAS, or other agencies mentioned herein.

CORRESPONDENCE

Please send all correspondence to kristin.a.ludwig@gmail.com.

Executive Summary

Large scale multi-user research infrastructure is a critical component of the federal science and engineering research enterprise. Developing infrastructure for multidisciplinary research requires large investments over long periods of time and typically involves partnerships among multiple geographically-distributed institutions for planning and implementation. In the scientific, political, and public spheres, large multi-user research facilities are high visibility endeavors, and engage thousands of domestic and international stakeholders across academia, industry, government, and the public.

The National Science Foundation (NSF) supports large research facilities that are created in response to community need and that span a broad range of disciplines including physics, astronomy, materials research, geosciences, ecology, engineering, nanotechnology, and polar research. Collectively, NSF's facilities are characterized as shared-use infrastructure, instrumentation and equipment that are accessible to a broad community of researchers and/or educators. Each year, NSF-supported facilities provide access to thousands of researchers. In FY2011, NSF invested $1.1 billion dollars in large facilities (construction and operations costs) with the operating costs of its operational facilities ranging from $7 to 98 million per year.

Multiple policy questions surround federal investments in large research facilities: who is benefiting from these investments? What is the best way to maximize scientific productivity across the research enterprise? How should investments in big science be balanced with support for individual or small group research? Ultimately, decisions on these issues are directly related to and significantly impact the individuals who constitute the scientific research community. For NSF facilities, the answers to these questions become focused on the activities of individuals who are interacting with the facility for the purpose of furthering scientific research and/or education – the "users."

This study provides the first known analysis of facility utilization at NSF. Four NSF-supported large facilities are used as case studies to create a conceptual framework for characterizing facility utilization, to identify how facilities know who their users are and how they are using their facilities, to examine changes in use over time, and to define how lessons learned from user analysis can be applied to facility management and planning. Results show that there is a broad spectrum of users who interact with each facility in different ways and that NSF is likely serving many more users than previously thought. New users discover facilities through different mechanisms; for some facilities, unanticipated users are driving new areas of research. Facilities enabled by cyberinfrastructure are experiencing rapid increases in data use and in some cases, the next generation of large facility users appears to be developing new skills for working in an increasingly data-intensive research environment.

This study suggests that characterizing and quantifying large facility use will likely become increasingly important as the federal government continues to focus on developing metrics and evaluation tools for assessing its investments in science and engineering research. It shows that facility users and the type of use may change as science and technology change over time and points to the importance of facilities recognizing the opportunities for growth and the need to balance these opportunities with their mission. Analyses in this study show that trends and observations in facility utilization across NSF can indicate areas of synergy and possible new avenues of collaboration between facilities, centers, and initiatives that may otherwise go unseen. Finally, changes in new user skills, backgrounds, and expectations may be important indicators of future needs for workforce development and user training. This work establishes a foundation for evaluating facility use and shows that this area is ripe for future work that may include portfolio-wide analyses, network or community mapping, and applications to other research investments such as mid-scale infrastructure or science centers.

1. INTRODUCTION

Across the science research enterprise, large-scale multi-user research facilities provide unique capabilities to a wide variety of scientific disciplines. These facilities require large investments (on the order of billions of dollars) over long periods of time (decades) and typically involve partnerships among multiple geographically-distributed institutions for planning and implementation. In the scientific, political, and public spheres, large multi-user research facilities demand special attention because they are high visibility endeavors, and engage thousands of domestic and international stakeholders across academia, industry, government and the public.

In the US, several federal agencies support large scale multi-user research facilities, including the National Science Foundation (NSF), the Department of Energy (DOE), and the National Aeronautics and Space Administration (NASA). Each agency follows its own protocols for planning, construction, operation, and management of its facilities. This report focuses on "large" facilities supported by NSF.

1.1 NSF's Large Facilities

The National Science Foundation (NSF) supports large research facilities that span a broad range of analytical and data collection capabilities such as mobile platforms, large experiments, and lab and field instrumentation. Facility research disciplines include physics, astronomy, materials research, geosciences, ecology, engineering, nanotechnology, and polar studies. Examples of current NSF-supported facilities include optical and radio telescopes for astronomy, global networks of environmental sensors for the earth sciences, ships for oceanography, and magnet labs and gravity wave detectors for physics research.

NSF's facilities are characterized as shared-use infrastructure, instrumentation and equipment that are accessible to a broad community of researchers and/or educators. The budgetary definition of a NSF "large" facility is one that (a) requires over $8 million dollars per year in operating costs (NSF FY2011 Budget) and/or (b) is constructed using funds from the Major Research Equipment and Facility Construction (MREFC) account.[1] Currently, NSF has 19 facilities in operation, six in construction, and more in planning (Appendix A). Each year, NSF-supported facilities provide access to thousands of researchers. In FY2011, NSF invested $1.1 billion dollars in large facilities (construction and operations costs) with the operating costs of its operational facilities ranging from $7 to 97 million per year.[2]

At NSF, each facility is developed in response to community need to tackle complex questions. Accordingly, NSF's large facilities are created through a complex process that begins *organically*, with informal and eventually formal discussions amongst community members. During this horizon planning phase, research communities work together through community workshops and focused working groups to identify research questions and the infrastructure needed to accomplish these goals.[3] Once these needs are articulated, the project graduates to a conceptual design phase, then proceeds through several mile markers of preliminary design and final design approval processes before construction begins.[4] Only construction costs are supported by the MREFC account; funding for both the conceptualization/planning *and* operation phases originates from the lead directorate's Research and Related Activities (R&RA) account. In its entirety, the process of NSF facility conception, development, and construction can take

[1] The NSF Large Facilities Manual describes the history of the MREFC account and project eligibility for MREFC funding.
[2] NSF FY2011 Budget (actual expenditures, reported in the FY2013 Request to Congress). This range excludes the US Antarctic Program, which had an operating cost of $297 in FY11.
[3] See "Appendix C - Histories of Projects Funded by the NSF MREFC Account" of the 2005 NRC Report *Setting Priorities for Large Research Facility Projects supported by the NSF* for multiple examples.
[4] For MREFC projects, this process is detailed in the NSF Large Facilities Manual.

decades to complete.[5] It is important to note that NSF does not construct, manage, or operate the facilities it supports: these responsibilities are awarded to external organizations ("awardee" institutions) such as universities, consortia, or non-profit organizations. The awardee institutions therefore play a critical role in interfacing with funding agencies, facility users, and the research community at large.

1.2 Motivation for Studying Facility Utilization

Analysis of users and large facility utilization can benefit funding agencies, the scientific research community, and facilities by drawing connections between research investments, research activity and scientific productivity. In turn, this information can ultimately inform funding decisions, community research priorities, workforce development needs, data management requirements,[6] facility lifecycle planning, international and commercial partnerships, STEM education initiatives, and performance evaluation. While each facility serves a diverse set of users with variable backgrounds and needs, NSF facilities share many challenges - such as data management and storage, and adapting to evolving technology and areas of research - many of which are connected to facility use. Recognizing these connections to user community, studying facility utilization has multiple benefits:

- **Assessing the vitality of a facility:** information on how users are interacting with a facility over time can be used as one measure of determining the value and productivity of the facility throughout its lifecycle.

- **Identifying areas of synergy:** NSF facilities are under increasing pressure to maximize partnering with other agencies and across NSF directorates.[7] Understanding facility use can be helpful for highlighting existing partnerships and for identifying new avenues for collaboration.

- **Informing facility management and planning:** Utilization information may facilitate program management and agency planning by identifying areas of cross-fertilization between fields and/or changes in community use.

- **Improving transparency to resource allocation:** Agency annual and strategic budgeting focus on construction and operation costs and do not provide much information on how resources are allocated. Information on facility use, paired with resource allocation, can be used to inform metrics on return on investment.

- **Optimizing user support:** By better understanding user needs and their interactions with the facility, user support can be improved to maximize scientific return on investment.

This study aims to characterize who is using NSF's large facilities, the metrics by which usage is assessed, and how anticipated users compare to actual users once the facility is in operation. To the author's knowledge, no studies on facility utilization across NSF have been completed.

[5] National Research Council, 2005
[6] At the April 2012 Annual NSF Large Facilities Workshop, the subject of data management (including archiving and curation) was the focus of many discussions. Participants agreed that planning for data management has implications across many aspects of planning and operating a facility including lifecycle planning, workforce development, and facility access. This will likely be a special topic for the 2013 workshop.
[7] America COMPETES Reauthorization Act of 2010, Sec. 523

1.3 Research Questions and Methods

Research Questions
To characterize the utilization of NSF's large facilities, this study focuses on three questions:

1. What is the definition of a NSF large facility "user?"

2. Who is using NSF's large facilities and how are they using these facilities? How does facility use change over time?

3. How can lessons learned be applied to managing existing facilities and planning for future investments?

Methods: Overview
This study employed multiple approaches to answering the research questions. Case study methodology was used to select and investigate a sampling of representative facilities. Qualitative data was collected through the observation of facility reviews and meetings; unstructured interviews; and site visits. Quantitative data on facility utilization was collected from the review of relevant reports, databases, and facility documents. These methods are described below and further details are provided in Appendix B.

Although this work focuses on observations from four case study facilities, other facilities in construction or planning phases (e.g., the Atacama Large Millimeter/submillimeter Array (ALMA), the National Ecological Observatory Network (NEON)) were examined in brief to investigate planning for facility use during the development phase of a facility. In addition, other facilities in operation (e.g., the Integrated Ocean Drilling Program (IODP), the National Nanotechnology Infrastructure Network (NNIN)) were not examined in depth, but the author's observations are included in the report where available and relevant.

To understand utilization of multi-user research facilities *outside* NSF, information provided by facility managers and senior administrators outside NSF was collected from a select set of interviewees. These included representatives from the Space Telescope Science Institute (STScI), which is supported by NASA and manages the Hubble Space Telescope (HST); DOE's Office of Science Basic Energy Sciences (BES); and the National User Facility Organization (NUFO).

Early in the project, feedback on study design and approach was gathered from meetings with analysts at the Science and Technology Policy Institute (STPI) and the White House Office of Science and Technology Policy (OSTP).

Methods: Case Study Selection
Because of the breadth of NSF's large facility portfolio (Appendix A), a case-study approach was used to structure this analysis, where each facility constituted one "case." Several constraints were emplaced at the outset of the case selection process: 1) selected cases were limited to facilities in full operation (i.e., excluding those that are in construction or in planning) because of the need to examine *actual* users; 2) Antarctic facilities (IceCube, the US Antarctic Program (USAP)) were excluded because of challenging access for conducting site visits.

To select cases, operational facilities were grouped by categories including primary research discipline, geographic location, and facility age (Table B-1). The following six factors were used to achieve a balanced set of cases for comparison: 1) facility age (marked by when the facility began operations); 2) primary research discipline; 3) the presence of a cyberinfrastructure (CI)[8]; 4) the physical facility setting (whether the facility occupies a single site (e.g., one telescope on a mountain), a group of distributed sites (e.g., labs at multiple universities), or a distributed network (e.g., a network of sensors in the natural environment); 5) location of facility operations (domestic vs. international); and 6) user access (whether users access the facility physically or remotely) (Table 1). As a result, four facilities were selected:

1. **Academic Research Fleet (ARF)**: The ARF consists of 21 ocean research vessels in the University-National Oceanographic Laboratory System (UNOLS). Vessels range in size, endurance, and capabilities, enabling NSF- and other federally-funded scientists with the means to conduct ocean science research with a diverse fleet capable of operating in coastal and open ocean waters around the globe. This study did not include human occupied vehicles or remotely operated vehicles.
2. **EarthScope:** EarthScope is a distributed, multi-purpose geophysical instrument array that helps researchers make major advances in our knowledge and understanding of the structure and dynamics of the North American continent. EarthScope is cyber-infrastructure enabled and is comprised of three parts: the USArray, the Plate Boundary Observatory (PBO), and the San Andreas Fault Observatory at Depth (SAFOD).
3. **George E. Brown Network for Earthquake Engineering Simulation (NEES):** NEES is a national, networked simulation resource of 14 advanced, geographically distributed, multi-user earthquake engineering research experimental facilities with telepresence capabilities. Experimental facilities include shake tables, geotechnical centrifuges, a tsunami wave basin, large-scale laboratory experimentation systems, and mobile and permanently installed field equipment.
4. **National Optical Astronomy Observatory's Kitt Peak National Observatory (NOAO/KPNO)**: NOAO is a Federally Funded Research and Development Center (FFRDC) for research in ground-based, nighttime, optical, and infrared (O/IR) astronomy. KPNO operates three major nighttime telescopes (supported by NSF) and hosts the facilities of consortia which operate 22 optical telescopes and two radio telescopes on the summit of Kitt Peak.

The similarities and differences between these facilities with respect to the six selection factors are illustrated in Table 1. Because of the differences in NSF's large facilities (Appendix A, Table B-1), it is difficult to select any one facility that is fully *representative* of the entire portfolio. However, case study methodology supports intentional sampling and lends itself to comparative analysis and proved to be a useful tool for this study.

[8] The term "cyberinfrastructure" has different meanings across NSF and its facilities: here it is used broadly. One definition, from an NSF workshop, is "the coordinated aggregate of software, hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering" (Berman, 2005). Another definition is "a distributed information technology infrastructure comprised of systems, software, databases, and visualization facilities, all interconnected with high-speed networks" (Hacker et al, 2012).

Table 1: Selection Criteria for Case Study Facilities

Criteria	ARF	EarthScope	NEES	NOAO/KPNO
Year Began Operation	1971	2004	2004	1960s
Primary Discipline	Oceanography	Geology	Engineering	Astronomy
CI-Enabled?	No*	Yes	Yes	No
Physical Site Type	Distributed Single Sites	Single Site & Distributed Network	Distributed Single Sites	Single Site
Operation Locations	Domestic & International	Domestic & International	Domestic	Domestic
User Access	Physical (remote access to select archived data)	Physical & Remote	Physical & Remote	Physical (remote access to archived data)
FY11 Operating Budget (dollars in millions)	$80.00	$25.05	$22.00	$31.50**

* Some vessels in the UNOLS fleet support participation of shore-based scientists and the public through the use of telepresence. Visual data is streamed online in real time for select expeditions.
**Budget for all of NOAO (includes KPNO and Cerro Tololo)

Methods: Qualitative Data Collection

During the course of this study, several meetings coordinated by or in conjunction with the Large Facilities Office provided useful insight into facility management, oversight, and use. These meetings (Table B-2) included Operations and Maintenance Reviews, Business Systems Reviews, and a special meeting on facility recompetition. Observations from these meetings informed the outcomes of this report.

Site visits to case study facilities were used to conduct interviews with facility staff (Table B-3). Interviewees represent a sampling of facility staff and include senior administrators (e.g., facility directors, assistant directors; project managers); IT managers and staff; engineers; site operations managers, education and public outreach specialists; and researchers (staff scientists, academic scientists, user committee representatives). A total of 78 interviews were completed for this work (Table B-4).

Methods: Quantitative Data Collection

Because each facility is unique, a variety of materials was used to collect quantitative data on each case study facility. These included: quarterly and annual reports; data from user-access of online databases (e.g., Google Analytics, user registration information); user surveys; user committee reports; facility and awardee institution websites; reports to NSF; NSF budget request to Congress; and National Academies reports. Many of these resources are not in the public domain and were obtained directly from program managers and facility staff. Any content directly pulled from these sources is presented in aggregate form in this report unless otherwise noted.

2. A Conceptual Framework for Characterizing Facility Utilization

2.1 Defining Large Facility "Users"

Many federal agencies support multi-user research facilities including NSF, DOE, and NASA. Across the US government, there is no single definition of a research facility "user:" the definitions are quite variable both across and within agencies due to the variability in facility types, access, and goals. Each agency has its own stakeholders, responsibilities, practices, culture, and history: no general attempt has yet been made to devise a federal-wide definition. In general, where an explicit definition exists, it is used to guide the scope of facility usage metrics (this is particularly true with DOE and is described further in Chapter 4).

Below, a working definition for NSF Large Facility users is derived from information gathered in this study. To illustrate the variability in user definitions across agencies and other entities, several examples from DOE, NUFO, STScI, and a National Research Council (NRC) report are included as a point of comparison.

National Science Foundation (NSF)
Like most agencies, NSF does not have an agency-wide definition of a "user" of its large research facilities. This is likely a reflection of the fact that NSF's facilities are community- rather than mission-driven; and, each NSF facility is unique.[9]

However, to create a framework for discussing utilization, it is necessary to formulate a definition of a facility "user." To this end, interviewees for this study were asked to provide a definition of their facility's users. Most sources initially cited examples of some type of researcher, student, or, in some cases, educator, *directly interacting* with facility resources such as data, data products, equipment, and educational materials. When asked to provide an explicit definition of a user at their facility, the answers ranged broadly. Some examples are below.

A user is/users are…

- "someone who has funding to use my lab"
- "people who don't have access to their own telescope"
- "people doing longer term projects"
- "someone who takes information from our facility to use it for scientific research and educational purposes"
- "someone who writes a proposal who wants to carry out work on one or more telescopes…"
- "someone to whom we provide access to telescopes and ensure quality data"
- "someone who interacts with the facility"
- "people who had traditionally done GPS work in the EarthScope footprint who saw value in this scale of work."
- "scientists, students, and sometimes the general public"
- "[someone] who uses the telescopes or archives for science; who uses our staff for technical consultation; who provides technical reviews of proposals. [Users include] public, students, and amateur astronomers."
- "a grad student/post-doc/faculty member/educator who uses [our facility's] data for geoscience research and education."
- "a researcher interested in testing at a large scale something to mitigate earthquake or tsunami risk"
- "someone who has been assigned or allocated time on the telescopes"

---

[9] In general with NSF facilities, it is uncommon to find a "one-size-fits-all" definition for terms relevant to facilities operations and management.

The responses clearly point to the intended users of NSF's facilities (a "broad community of researchers and/or educators"), but also show that there are different types of users and that users access and interact with NSF facilities in different ways, which is addressed later in this report. As a conclusion, a working definition of a NSF facility user is:

> **A <u>user</u> of a NSF large facility is someone who interacts with the facility for the purpose of furthering scientific research and/or science education in both formal and informal environments. A user of a NSF research facility is not necessarily supported by NSF funding.**

This definition is deliberately broad to reflect the research and educational mission of NSF[10] and is adopted going forward in this report.

Department of Energy (DOE)
The DOE Office of Science Basic Energy Sciences (BES), which supports research at light sources, neutron sources, nanoscale science research centers, and electron beam microcharacterization centers,[11] has a very prescribed definition of its users:[12]

> "Users are researchers who propose and conduct peer-reviewed experiments at a scientific facility. There are two other types of users who conduct experiments: (1) Remote User—a researcher who has been granted authority to remotely produce data (this excludes persons who can "look at data"); and (2) Off-Site User—a researcher to whom the facility provides custom-manufactured materials, tools, or devices that the facility has unique or unusual capabilities to fabricate (this applies only to such activities at Nanoscale Science Research Centers). For both types of these users, only one user is to be counted per proposal regardless of the number of co-investigators, and only if no individual is counted in any of the other user categories under the same proposal… Users must submit a successful, peer-reviewed research proposal and conduct experiments, as described above. Therefore, users do not include individuals who only send in samples to be analyzed, even if such activities are part of a peer-reviewed experiment. Users do not include individuals who pay to have specialty services performed or visit the facility for tours or educational purposes. Users also do not include researchers who collaborate on the proposal or subsequent research papers but do not conduct experiments at the facility."[13]

By comparison, DOE's Atmospheric Radiation Measurement Climate Research Facility (ARM), supported by the DOE Biological and Environmental Research (BER) division, provides resources to users who physically visit ARM and those who wish to download archived data from ARM's climate database. It defines its users as "site visitors as well as computer accounts of those requesting data."[14]

---

[10] NSF was created by Congress in 1950 "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes." NSF is committed to supporting research and education at all levels across a broad range of science and engineering disciplines: this goal is described in detail in the NSF Strategic Plan FY2011-2016.
[11] See DOE's Current List of Office of Science User Facilities at http://science.energy.gov/~/media/_/pdf/user-facilities/Office_of_Science_User_Facilities_FY_2012_rev1.pdf
[12] DOE National Facility users are often called "badged users" because of the security badge they have to have to access and use the facility.
[13] DOE Office of Science Scientific User Facilities Division website, http://science.energy.gov/bes/suf
[14] See http://www.arm.gov/about/stats

National User Facilities Organization (NUFO)

The National User Facilities Organization (NUFO, www.nufo.org) "represents the interests of all users who conduct research at US national scientific user facilities, as well as scientists from US universities, laboratories, and industry who use facilities outside the United States" and has 46 facility members,[15] including seven NSF large facilities (CHESS, LHC, Arecibo, NHMFL, NOAO, NRAO, and NSCL). NUFO does not have an explicit definition for a facility user and recognizes that its member facilities each have their own definition of their users. However, as of the June, 2012 Annual NUFO meeting, the organization is working towards creating a definition of facility users (see discussion in Chapter 7).

National Aeronautics and Space Administration (NASA)/Space Telescope Science Institute (STScI)

Staff at the NASA-supported Space Telescope Science Institute (STScI), which manages the Hubble Space Telescope (HST), did not know of an "official" NASA definition of user and "expected that the definition is variable across NASA's research facilities" (e.g., the HST, Kepler, and the future James Webb Space Telescope). According to one source, a user is "someone interested in applying for HST time or someone who has been awarded HST time" and/or "someone who accesses and uses archived and publicly available data." Another interviewee defined a user as "anyone who goes to the archive or anyone who can generate new ideas and theories from the data."

National Research Council (NRC)

In a 1999 National Research Council report on facilities dedicated to research in synchrotron radiation, neutrons, and high magnetic fields, "users" are explicitly defined as "on-site researchers who conduct experiments at facilities" where "an individual is counted as one user (per facility annually) regardless of the number of visits in a year." A "typical" user of the high magnetic field labs is described as:

> "a member of a small research group based in an academic institution, a national laboratory, a for-profit corporation, the facility itself, or a similar foreign institution that is supported by individual investigatory grants from agencies like NSF, NIH, and DOE, or by corporate fund. The user generally visits the facility a few times a year to collect data that cannot be obtained using ordinary laboratory equipment. The users have varying levels of experience with the technologies at these facilities… ." [16]

Clearly, the definitions of users vary both across and within agencies. For this reason, it is important to establish a definition of a facility user as the foundation for analyzing facility utilization.

2.2 Types of NSF Large Facility Users

In general, there are many kinds or "types" of users across NSF's facilities and each type interacts with the facility in different ways. Across the case study facilities, several common user types emerged from the interviews, site visits, and meeting observations completed during this study. As a result, seven user types are defined below, with examples from the case study facilities (Table 2). It is important to recognize that individual users may be one or more "type" of user. For example, a professor who is a "scientist" can also be an "educator" while teaching her graduate students or undergraduates about her work.

[15] See http://nufo.org/facilities.aspx

[16] National Research Council, 1999

Table 2: Case Study Facilities and Types of Users

Facility	Investigators (incl. post-docs)	Grad. Students	Undergrad. Students	Educators*	Comm/ Industry Reps	Citizen Scientists/ Amateurs	Public*
ARF	✓	✓	✓	✓			✓
EarthScope	✓	✓	✓	✓	✓		✓
NEES	✓	✓	✓	✓	✓		✓
NOAO/KPNO	✓	✓	✓	✓		✓	✓

** Not all interviewees agreed that educators and the public are "users." See discussion below for special notes on these two user types.*

Investigators

An *investigator* is someone who works as a professional researcher and conducts and/or contributes to peer-reviewed scientific research. Here, a post-doctoral researcher ("post-doc") is an investigator. This type includes investigators at both academic and federal institutions and professional research scientists at the facilities. Across the case study facilities, investigators (termed as "scientists," "researchers," "PIs," "co-PIs," "faculty," and "professional scientists" by interviewees) and post-doctoral researchers were always cited as users of a facility. (In the case of NEES, research engineers are a type of investigator.)

Graduate Students

At facilities, a *graduate student* is someone studying science and conducting original research by accessing facility instrumentation, equipment, and/or data, typically in a field related to his/her academic advisor(s). This category includes students pursuing both master's and doctoral degrees in science. Across the case studies, graduate students were always included as users of the facility.

Undergraduate Students

An *undergraduate student* is a student pursuing a bachelor's degree or foreign equivalent. Case study facilities that support undergraduate research through either formal or informal programs cited undergraduate students as facility users. NEES, EarthScope, and NOAO/KPNO all participate in the NSF-sponsored Research Experience for Undergraduate (REU) program[17] and this is the most common path for undergraduate students to engage with these facilities. Vessels in the ARF sometimes have REU students sailing onboard where the REU program is arranged through either the operating institution or another university. In other cases, the operator receives state or other funding to support several ship days per year of dedicated undergraduate student use. In all of these cases, undergraduate students are considered users of the Fleet.

Commercial/Industry Representatives

A *commercial/industry representative* is someone who works in a commercial, non-academic, or for-profit company or organization. Of the four case study facilities, only EarthScope and NEES cited example users of this type. At EarthScope's Plate Boundary Observatory (PBO), surveyors and representatives from public utilities companies (e.g., Seattle Public Utilities) have used GPS data for planning power lines, pipelines, water distribution systems, and building monitoring. At NEES, earthquake engineering practitioners and building material manufacturers use data and results from NEES experiments for informing building code and construction projects.

Educators (Formal and Informal)

An *educator* is a professional teacher, instructor, or educator who works in either a formal or informal educational environment at any level of education (e.g., formal education in grade school through higher

[17] See http://www.nsf.gov/crssprgm/reu/

education; informal education at museums, learning centers, after-school programs, and cyber-learning). All of the case study facilities support education and public outreach programs designed to share and promote the scientific research they support[18] and these programs vary greatly in scope both within and across NSF's large facilities. Opportunities for educators to "use" the four case study facilities include participating in field campaigns, attending professional development programs, and using facility-produced classroom or curricular materials.[19]

Citizen Scientists/Amateurs

A *citizen scientist* is an amateur or non-professional scientist who collects data, typically in the form of observations or samples, and contributes these data to established research programs. In many disciplines, "citizen science" programs have increased over the past several years and are increasingly taking advantage of the Internet and mobile devices to engage non-professional scientists of all ages and backgrounds in some aspect of the process of science.[20]

Of the four case study sites, only NOAO/KPNO definitively included amateur astronomers and/or citizen scientists in their list of users. For example, leaders of the popular Nightly Observing Programs at Kitt Peak[21] are able to use the educational telescopes (research telescopes that have been re-purposed for educational use) for personal use after the tours.[22] A second example includes activities during the International Astronomy Year, where KPNO education staff created "make your own telescope" kits ("Galileoscopes") and held multiple public events at the summit of Kitt Peak to enable amateur astronomers to use Kitt Peak's geographic location and staff expertise for personal observations.

By comparison, even though anyone can download EarthScope data, the degree to which these data are used by citizen scientists is difficult to determine. While EarthScope's USArray and PBO notice "spikes" in website visits in the aftermath of seismic events such as the August, 2011 earthquake in Virginia,[23] these visits are most likely from the curious public and research scientists than from citizen scientists. With ARF and NEES, citizen scientists were not included as users. The involvement of this user type at these facilities is likely limited by the need to physically access the ships and testing labs.

The Public

A *public* user is any member of the general public who interacts with the facility by participating in facility-sponsored public events and resources such as tours, open houses, and visitor centers. The

[18] These programs play an important role in the fulfillment of NSF's broader impacts requirements.
[19] Curricular materials are produced by EarthScope, NEES, and NOAO/KPNO but not ARF.
[20] There are many examples of successful citizen science programs in different disciplines. The astronomy community has a rich history of "amateur" participation in astronomical observations and classifications (e.g., http://www.galaxyzoo.org/). Ecologists have engaged citizens in plant bloom surveys to examine long term and seasonal trends over large geographic scales (e.g., NEON's Project BudBurst, http://neoninc.org/budburst/). The US Geological Survey's (USGS) popular "Did You Feel It" website (http://earthquake.usgs.gov/earthquakes/dyfi/) invites citizens to record earthquakes that they experience and these data are used to inform the location and impact of seismic events. New websites such as SciStarter (http://scistarter.com/) function like a clearinghouse for citizen science opportunities.
[21] See http://www.noao.edu/noao/pio/pop/
[22] The NSF-supported 2.1m, Mayall, and WIYN telescopes at NOAO/KPNO are not available for educational use.
[23] For example, see the seismic data visualization developed by IRIS at http://youtu.be/IKE7MLNdtcg and community discussions on GPS data on UNAVCO's website at http://www.unavco.org/voce/viewforum.php?f=55.

opportunities for the public to use NSF facilities vary greatly by facility.[24]

All of the case study facilities either regularly or occasionally offer formal and informal tours to the public, school groups, and interested professionals.[25] All of the facilities have participated in either "open house" or large public events like the USA Science and Engineering Festival.[26]

Of the four case study facilities, NOAO/KPNO has the highest degree of public involvement: their visitor center sees thousands of visitors per year.[27] The public can also "experience" NOAO/KPNO through their popular Nightly Observing Programs. By comparison, vessels in the ARF will occasionally have open houses and public tours of the ship (In 2006, the University of Washington had >800 visitors tour the R/V *Thomas G. Thompson* over two days). Similarly, many of the NEES facilities provide public tours of their test facilities. EarthScope, because of its distributed nature, does not provide tours, but it does have a vibrant education and public outreach program that even includes kiosks at highway rest stops.

"Other" Users

Depending on the facility, individuals who do not fit into any of the above categories were cited as examples of "other" facility users. The most common type of "other" users described were particular members of *Administrative or Support Staff,*[28] individuals who provide technical expertise to supporting research and/or education completed using facility resources. For example, EarthScope PBO data managers and KPNO engineers use facility-generated data or facility-housed equipment to improve data products or develop new instruments, respectively. At NOAO/KPNO, a few support staff are also "investigator" type users (they maintain a small research program in addition to their administrative duties: they compete for telescope time and are not given any kind of priority access). Due to the variability in "other" users, this was not included as one of the primary categories of facility user types.

International Users

Not all users of NSF's facilities are US citizens or affiliated with a US research institution and all of the case study facilities have some level of international use. NEES users can have NEESR or other NSF grants, or be supported by other sources, including the private sector.[29] Anyone from anywhere can freely download data from the NEESHub.[30] By comparison, NOAO/KPNO awards time on its telescopes entirely

[24] An in-depth analysis of education and public outreach programs at each facility is beyond the scope of this study; however, there are many opportunities for comparative analyses of education programs across NSF's facilities.

[25] These activities are typically coordinated by the facility's education, outreach, and communications staff. However, some facilities (e.g., the NEES MAST lab) do not have a dedicated staff for these types of efforts and rely entirely on community volunteers to donate their time for the organization and execution of public outreach activities.

[26] See http://www.usasciencefestival.org

[27] In FY2011, the Kitt Peak Visitor's Center reported 16,423 visitors (this includes tours, school groups, and participants in the Nightly Observation Program) (NOAO FY2011 Annual Report). Since FY10, the Visitor Center and its exhibits and activities are no longer funded by NSF and they are primarily supported by revenue from visitor programs and retail shop sales.

[28] Not all support staff members are users. For example at the NEES MAST lab, the project managers who work with visiting scientists to conduct their on-site experiments do not consider themselves to be users. Similarly, most administrators and IT managers at all of the case study sites considered themselves to be support staff, but not users.

[29] Grantees not supported by NSF must pay appropriate facility user fees.

[30] Monthly updates to NEESHub usage metrics are available at http://nees.org/usage. For example in FY2011, only 38% of the total users (the sum of registered users, unregistered interactive users and unregistered download users; see website for explicit definitions of each user category) were from the US. Others hailed from Asia (33%), Europe (19%) and other (10%) countries.

by merit: in FY2011, 70% of its investigators (PIs and co-PIs) were from the US.[31] EarthScope users are also able to freely download data from anywhere and there are many international users (see Table 4). Scientists onboard the ARF vessels are not always US scientists; and users of the Rolling Deck to Repository (R2R) online data repository are located globally.[32] While the similarities and differences between domestic and international users are not the focus of this study, international use is best examined with user-tracking data, as discussed in Section 4.3. Ultimately, despite their geographic location, international users fall into one of the seven types of users described above.

DOE User Types

Administrators at DOE's Office of Science Basic Energy Sciences (BES) cited scientists, post-docs, graduate students, and "occasionally undergraduates" as their users. They did not consider educators or members of the public (e.g., tour participants) to be users of their facilities and emphasized that DOE's mission does not include education outside training graduate students, post-docs, and undergraduates. Many of DOE's BES Scientific User Facilities are also used by industry/the commercial sector for proprietary use.

NASA/HST User Types

HST staff at STScI described users as professional scientists, grad students, and post-docs. They included public citizens, amateur astronomers, citizen scientists, and artists as users of the Hubble and its imagery. The HST has multiple examples of citizen science/amateur and public use at a much broader and publicly visible scale than any of the NSF case study facilities.

Summary

While most of the user types identified above are not surprising, the most interesting result of this characterization is that **not all of the NSF case study facility interviewees consider *educators* and/or the *public* to be facility users**. At the case study sites, facility staff responses varied with respect to whether or not educators are included as users. Many agreed educators *are* users, but some interviewees held that educators are *not* users because they are not using the facility for scientific research and are not publishing data or peer-reviewed publications. Others thought that educators weren't "users" *per se* – but were "some other category" that remained undefined. The interpretation of the public as a user varied greatly across the interviewees and this often corresponded directly with their position and responsibilities with the facility. Typically, education and public outreach personnel and many senior leadership staff members claimed the public to be a type of facility user; whereas science, engineering, IT, and administrative staff did not always view the public to be a user.

Because facilities are designed to serve a broad community of scientists and educators, and because NSF is committed to science education in addition to supporting fundamental research, educators and the public are included as user types in this study.

In conclusion, there are seven primary "types" of NSF facility users: investigators, graduate students, undergraduate students, educators, commercial/industry representatives, citizen scientists/amateurs, and the public. Not all facilities have all types of users and this is due to the differences in what resources or opportunities are provided by each facility to the different user types. Some individuals may be different types of users: for example, an investigator or graduate student may be an educator; or an educator may be a citizen scientist.

---

[31] Detailed demographic information on KPNO observers is presented in the FY2011 Annual Report, available by request from the NSF program manager for NOAO.

[32] See Section 3.1 and Figure 5.

2.3 User Access to Facilities

Users access the facilities via different pathways to accomplish their work. Based on the case study analysis, the means by which users *access* the facilities can be distilled into two pathways: physical and remote:

- ***Physical* access** takes place on site and/or with its physical resources such as instrumentation or equipment.
- ***Remote* access** is done from a distance, via the web, phone, telepresence, remote servers, or social media.

Depending on the facility and user needs, users access facilities either physically or remotely, or by a combination of both pathways.

Historically, large research facilities like astronomical observatories (e.g., NOAO/KPNO or the Observatories) and physics experimental facilities (e.g., many of the DOE beam lines) have required users to *physically* be on site to conduct their research. However, the Internet increasingly enables new pathways for remote access to facility resources. This combined with a commitment to open data policies[33] has changed how users access and interact with some facility resources and facility-produced data. Long-term changes in remote access and remote users are discussed in greater detail in Chapter 6.

2.4 User Interactions with Facilities

Facilities provide different capabilities, resources, and services to their users. From the perspective of investigators and graduate students, these resources include data collection, specialized instrumentation, sample curation, testing platforms, equipment pools, consultation, training, data management, and data archiving (Figure 1). From the perspective of educators, these include curricular materials, teaching kits, training, equipment, and learning datasets. Different user types access different facility resources physically and/or remotely, depending on his or her needs (Figure 1). The frequency, duration, and “intensity” of a user’s interaction with a facility changes depending on the user.

For example, most users of NOAO/KPNO and vessels in the ARF schedule time to make direct observations using the Kitt Peak telescopes or to conduct expeditions at sea, respectively. In contrast, researchers who use different parts of EarthScope may never set foot in the field: some rely entirely on regular data downloads from EarthScope’s data streams. Others borrow instrumentation from EarthScope’s PBO and USArray to deploy in the field. NEES researchers and engineers schedule time to conduct experiments at one of the 14 NEES sites; other NEES users may observe an experiment using telepresence; and yet others contribute data and interact with colleagues online using the NEESHub “Project Warehouse.”

It is rare, especially for an investigator-type user, to follow only one path of use with the facility and the patterns of an individual’s interaction with the facility may change over time. For example, a relatively new investigator may initially *consult* with a facility to develop a new instrument or devise a new data analysis tool. The same individual may later become a regular “data miner,” where his/her facility interactions are primarily through remote access to a facility’s *data archives*. The investigator may physically attend a

[33] Beginning in January, 2011 all proposals submitted to NSF must include a “Data Management Plan” describing the PIs plan for sharing and disseminating results. See chapter II.C.2.j of the Grant Proposal Guide (http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp) and the associated press release (NSF PR 10-077) at http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928.

training workshop to develop new skills, and/or, may adopt facility *curricular materials* for classroom teaching.

In contrast, an investigator interacting with the same facility may only download data for one project. An educator or public user may attend only one event at the site of the facility. These are examples of "less intense" use by different user types who follow more singular pathways of interaction with the facility.

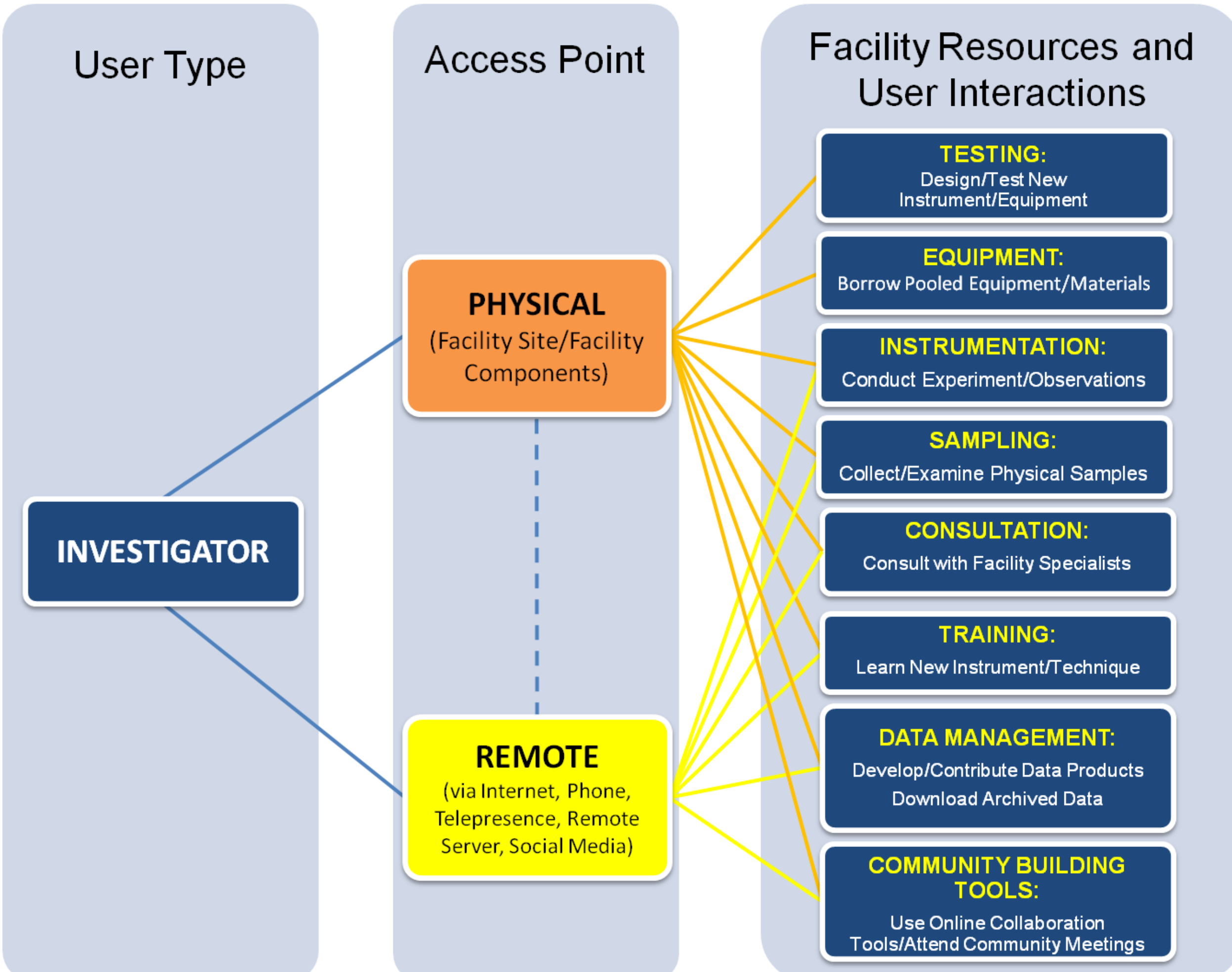


Figure 1: A schematic of how an investigator user accesses and interacts with a facility. *Facility services are indicated in yellow type. It is rare, especially for* investigator *user, to follow only one path of use with the facility.*

2.5 Facility Interactions with Users

When describing types of use, several interviewees at the case study sites described varying "degrees" or "levels" of users. For example, according to some, a PI who had a NSF grant would be considered a primary or first-level user, compared to a public citizen, who would be a lower-level user. Alternatively, some interviewees stated that users were only those who *directly interacted* with the facility as compared to others who take a tour or observe a demonstration. Other interviewees described "power users" – those who are regularly (e.g., daily or weekly) downloading new data and making heavy use of a facility's online tools versus other users who may be the equivalent of “window shoppers” and only experimenting with online resources before committing to using them on a regular basis.[34]

During discussions with interviewees, the determining factor for what made a user a “primary” or “top level” user vs. a secondary or tertiary user varied and typically fell into three categories: 1) whether or not the user had earned financial support to use the facility (e.g., a NSF or other grant); 2) the extent to which the user had a “hands-on” experience at the facility; and 3) the frequency or duration of use. These levels of users are illustrated in Figure 2, which differentiates user types into “tiers,” based on their level of interaction with the facility.

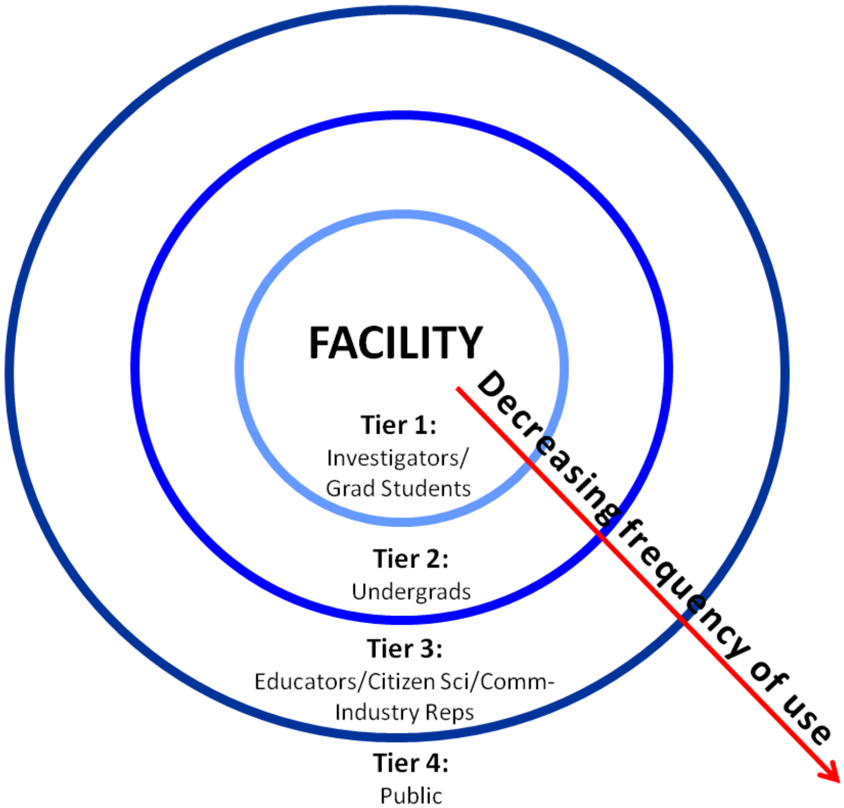


Figure 2: Schematic of User Tiers. *Different user types are sorted into four tiers, based on their level of interaction with the facility. At the core are investigators and graduate students, who are often called “power users.” The second, third and fourth tier users use facility resources to a lesser and lesser extent than the Tier 1 users.*

At the center are Tier 1 users, which are “core” users: investigators and graduate students who use multiple facility assets. Many Tier 1 users have directly received financial support or time to conduct research at the facility. Tier 1 users contribute to peer-reviewed literature. Tier 2 users are typically undergraduates or others who are in an “attendant” position relative to first level users. Tier 3 users are those who use facility products such as educational materials, or open source software for work that is not directly related to the facility. They are typically educators, citizen scientists, or commercial representatives. Tier 4 users interact with the facility to a lesser extent, such as attending a facility event. Tier 4 users are typically members of the public.

[34] See Hacker and Magana, 2011, who invoke “Roger’s Diffusion of Ideas” to describe user interactions with the NEES CI. “Window shoppers” are described as users who are exploring the facility resource, but have not yet fully adopted it for routine use in their research.

While creating user tiers is subjective, it provides a useful framework for characterizing how different user types interact with a facility and reflects the overall goals of a facility, which focus first on providing support to their research users, and invest fewer resources in supporting users at the outer tiers.

2.6 Discussion

This chapter establishes a conceptual framework for characterizing facility utilization to answer the first two research questions: what is the definition of a facility user? and, who is using NSF's large facilities and how are they using them? Results show that NSF facility users are defined as individuals who interact with the facility for the purpose of furthering scientific research and/or science education in both formal and informal environments. A user of a NSF research facility is not necessarily supported by NSF funding. There are seven types of NSF facility users, including investigators, graduate students, undergraduate students, educators, commercial/industry representatives, citizen scientists/amateurs, and the public, and not all facilities have all types of users. These different types of users interact with the facility in different ways, depending on their needs. User-facility interaction is categorized by the types of services that a facility provides to its user communities. For investigators, these resources include instrumentation, sampling, testing, equipment, consultation, training, data management, and tools for building and fostering a community. For every facility, there are multiple tiers of users where scientists and graduate students are the primary or "core" users. Other user types, such as undergraduates, educators, citizen scientists, and the public, fall into outer tiers, based on their frequency or intensity of facility use.

This framework of conceptualizing facility utilization sets the groundwork for investigating facility use at a more quantitative level, through user tracking. In the next chapter, multiple examples from the case study facilities show *how* facilities know who their users are and how their users are interacting with the resources they provide.

3. Analyzing Users and Facility Use through User Tracking

Facilities track their users to better understand who is using which assets of the facility. Depending on the type of facility and how users access the facility, different types of users are tracked differently. For example, some facilities track users who physically come to the facility site. Others track and examine trends in data downloads from remote users. In general, the goal of tracking users is to provide better "customer service" to the users by knowing what is heavily used and what could be improved to ultimately deliver more value to scientific and educational endeavors. User tracking can also be used to show the impact of a facility on a community (such as the degree to which the facility is being used or adopted by the community at large). Across the case study facilities, quantitative information on facility use is highly variable and is reported in different locations, such as quarterly and annual reports, science plans, and facility websites.[35]

The four case-study facilities use several kinds of user tracking methods including IP addresses, the number of users on site, the number of workshop attendees, the number of event visitors, and the number of refereed publications (Table 3). These methods primarily apply to Tier 1 users – or Investigators and Graduate Students. Other methods, such as the number of workshop or public event participants (which can apply to Tier 2, 3, and 4 users) are also used and are described below.

Table 3: Different User Tracking Methods Used by Case Study Facilities

User Tracking Method	IP Addresses*	# On Site	# Workshop Attendees	# Event Visitors	Refereed Publications
ARF	✓ (unreg.)	✓	✓		
EarthScope	✓ (unreg.)		✓	✓	✓
NEES	✓ (reg.)	✓	✓	✓	✓
NOAO/KPNO	✓ (reg.)	✓		✓	✓

* Some facilities require users to register to use their online resources. "Reg." indicates registration required; "unreg." indicates that registration is not required. See discussion below for examples of these tracking methods from the case study facilities.

3.1 IP Addresses, Data Downloads, and Website Visits

For facilities that are enabled by a cyberinfrastructure (CI) and that stream data live (e.g., EarthScope and NEES), and/or that have online data archives (e.g., EarthScope, NEES, NOAO's Science Archive, and ARF's R2R database), users can be tracked by Internet Protocol (IP) addresses. IP addresses can be used to resolve a) unique second level domains (e.g., .edu addresses indicate academic or research institutions; .gov addresses indicate government (federal, state, local) institutions; .org indicate non-profits; and .com indicate the private sector; international domains are indicated by country abbreviation, .tw indicates Taiwan, .nz indicates New Zealand, etc.); and, b) the geographic location of the computer from which the user is accessing facility data.

Example 1: EarthScope's Plate Boundary Observatory

In the case of EarthScope's Plate Boundary Observatory (PBO) data, anyone anywhere can access the data free of charge and without the need to complete any kind of registration. Using secondary IP

[35] Recognizing the heterogeneity of data on science programs and R&D investments is not new. The National Science and Technology Council (NSTC) 2008 report on "The Science of Science Policy: A Federal Research Roadmap" found that "agencies are using very different models, data, and tools to understand their investments in science and technology." This finding in part motivated the creation of the "science of science policy" field and the formation of what has become the STAR Metrics program, which is discussed in Section 6.1.

address tracking, PBO data managers can approximate the total number of users (Figure 3) and examine what types of PBO data are used the most heavily by different user types (Figure 4). The caveats with this type of tracking are that the same user may download data from different computers with different addresses; domain names can change over time; and multiple users can log in from the same computer or institution.

Some facilities (e.g., NEES, discussed below) require their users to register for online data access while others do not. During conversations with EarthScope personnel, several interviewees noted that EarthScope had originally intended to request user registration so that use could be better tracked and understood. However, pushback from the community was cited as the primary reason that registration for data is not required. Therefore, IP addresses are currently the best way to track the use of EarthScope PBO data.

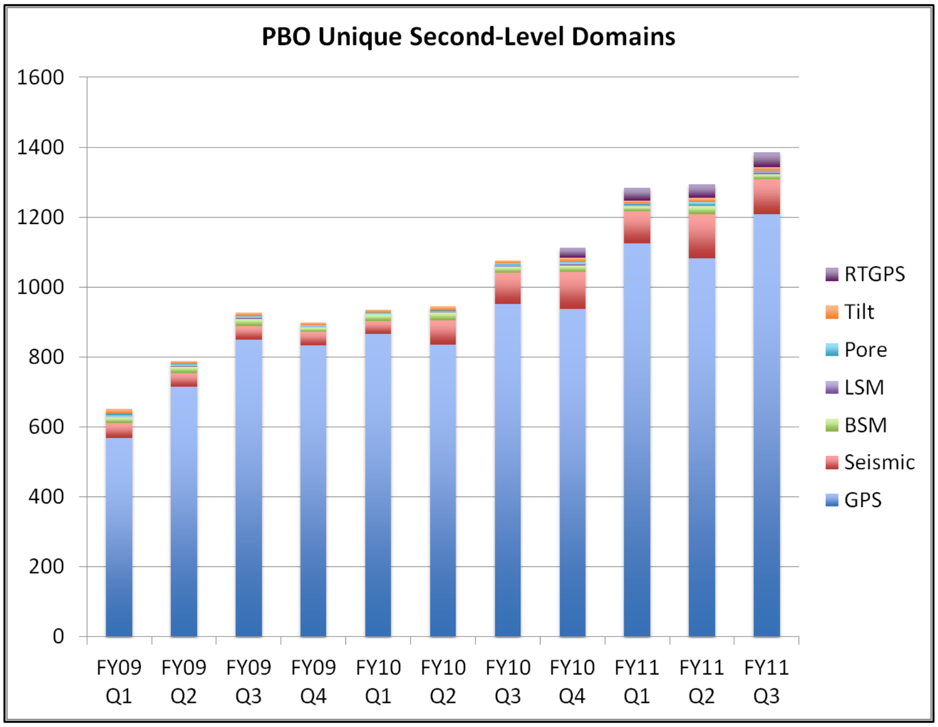


Figure 3: Approximating the number of users by second-level domains for EarthScope's PBO. *The number of unique second-level domains (y-axis) can be used to approximate the number of users using EarthScope PBO data (see legend for different types of data downloaded) over time (x-axis)). Clearly, since EarthScope's PBO data became available in 2009, the total number of users has steadily increased (data courtesy UNAVCO).*

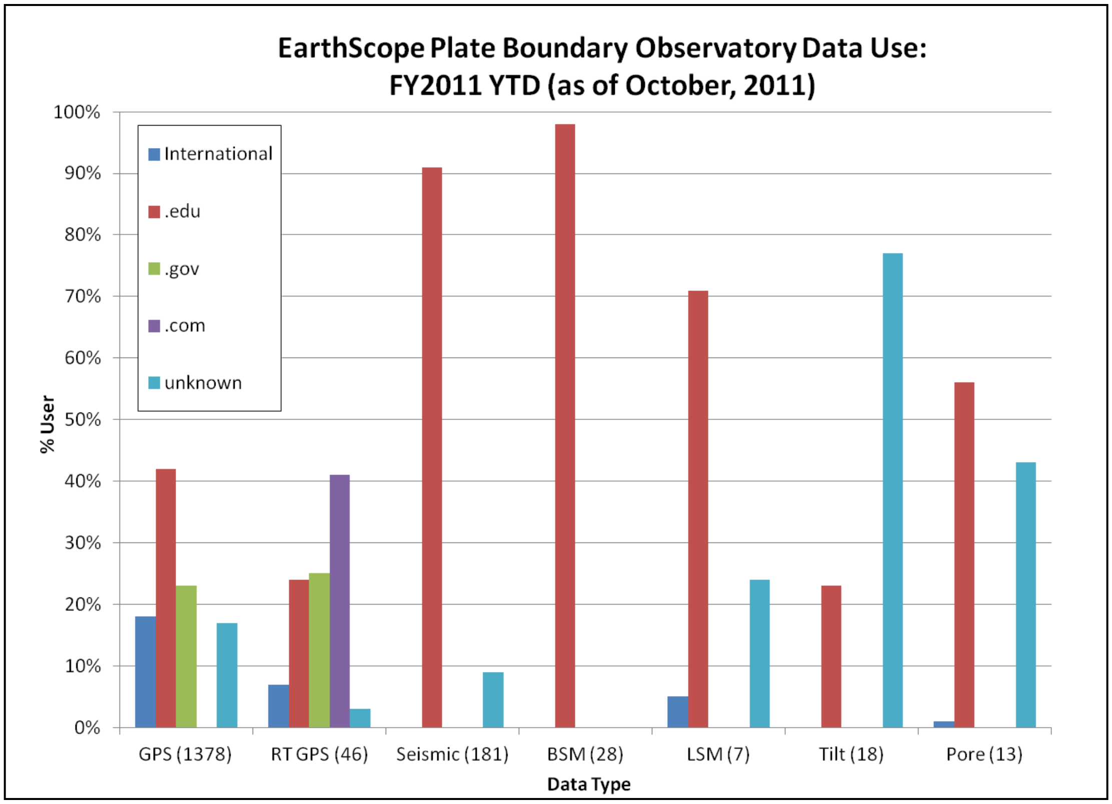


Figure 4: Types of Data Used by Different Types of Users of EarthScope's PBO. *EarthScope PBO Data Use (FY2011 through October, 2011) by data type (X-axis) and user type (y-axis), as estimated by secondary IP address (.edu users are likely scientists and/or students; .gov users are from federal/state/local government institutions; .com are from the private sector). On the x-axis, values in parentheses are the number of unique second-level domains (see Table 4). These data show that different types of users use different types of data: the majority of users are from academic (.edu) institutions; nearly 20% of the users of GPS data are international; and 40% of the users for Real Time GPS (RT GPS) data are from the private sector (data courtesy UNAVCO).*

Table 4: EarthScope PBO Users by Data Volume. *This table illustrates the "top users by data volume" off PBO data, as indicated by domain name (courtesy UNAVCO).*

FY2011 YTD	PBO							SAFOD		
User Statistics	GPS (1378)	RT GPS (46)	Seismic (181)	BSM (28)	LSM (7)	Tilt (18)	Pore (13)	Core	Seismic	Strain
Unique 2nd-level Domains	1378	46	181	28	7	18	13	16	0	0
international	18%	7%	<1%	<1%	5%	0%	1%	47%*	0	0
.edu	42%	24%	91%	98%	71%	23%	56%	47%*	0	0
.gov (federal + state)	23%	25%	0%	<1%	0%	0%	0%	6%*	0	0
.com	<1	41%	<1%	<1%	0%	0%	0%	0%	0	0
unknown	17%	3%	9%	<1%	24%	77%	43%	0%	0	0
Top Users	usgs.gov	Topcon	muohio.edu	ucsd.edu	nmt.edu	berkeley.edu	nmt.edu	USGS	none	none
by Data Volume	ncu.edu.tw	GPS Solution	ucsd.edu	berkeley.edu	ucsc.edu	pw.edu.pl	gc.ca	Utah State		
(does not include	cwu.edu	SIO	washington.e	princeton.edu	ucsd.edu	nmt.edu	gatech.edu	Penn State		
unknown domains	whoi.edu	CWU	uoregon.edu	ucsb.edu			ucsd.edu	Texas A&M		
or ISPs)	nagoya-u.go.j	Trimble	princeton.edu	usgs.gov				MARUM		
	utah.edu	gc.ca	mit.edu	gc.ca				U.J.F. Grenoble		
	nmt.edu	IGS	ucsb.edu					GFZ Potsdam		
	ucsd.edu	NOAA	sc.edu							
	berkeley.edu	USGS CVO	nanometrics.c							
	colorado.edu	AZGPS	usgs.gov							

* For SAFOD core, the user breakdown is by number of requesting PI institutions from Phase 3 Cycle 3 of sample requests

Although ARF and NOAO/KPNO are not CI-enabled, users are tracked by monitoring data downloads and website visits to their online data repositories.

Example 2: ARF's Rolling Deck to Repository (R2R) Online Data Storage

For the oceanographic community, NSF currently supports the development of the "Rolling Deck to Repository" (R2R) online data archive (http://www.rvdata.us) for underway data collected using vessels greater than 50' in length that receive NSF support or carry NSF investigators (e.g., the ARF and Sea Education Association vessels and coast guard cutters). Since R2R began in October, 2009, approximately 5,000-6,000 datasets from about 350 cruises are added to the archive each year. Data are free and currently users do not need to register or log in to download data. Information about users is collected using Google Analytics and can indicate information such as country of origin (Figure 5).

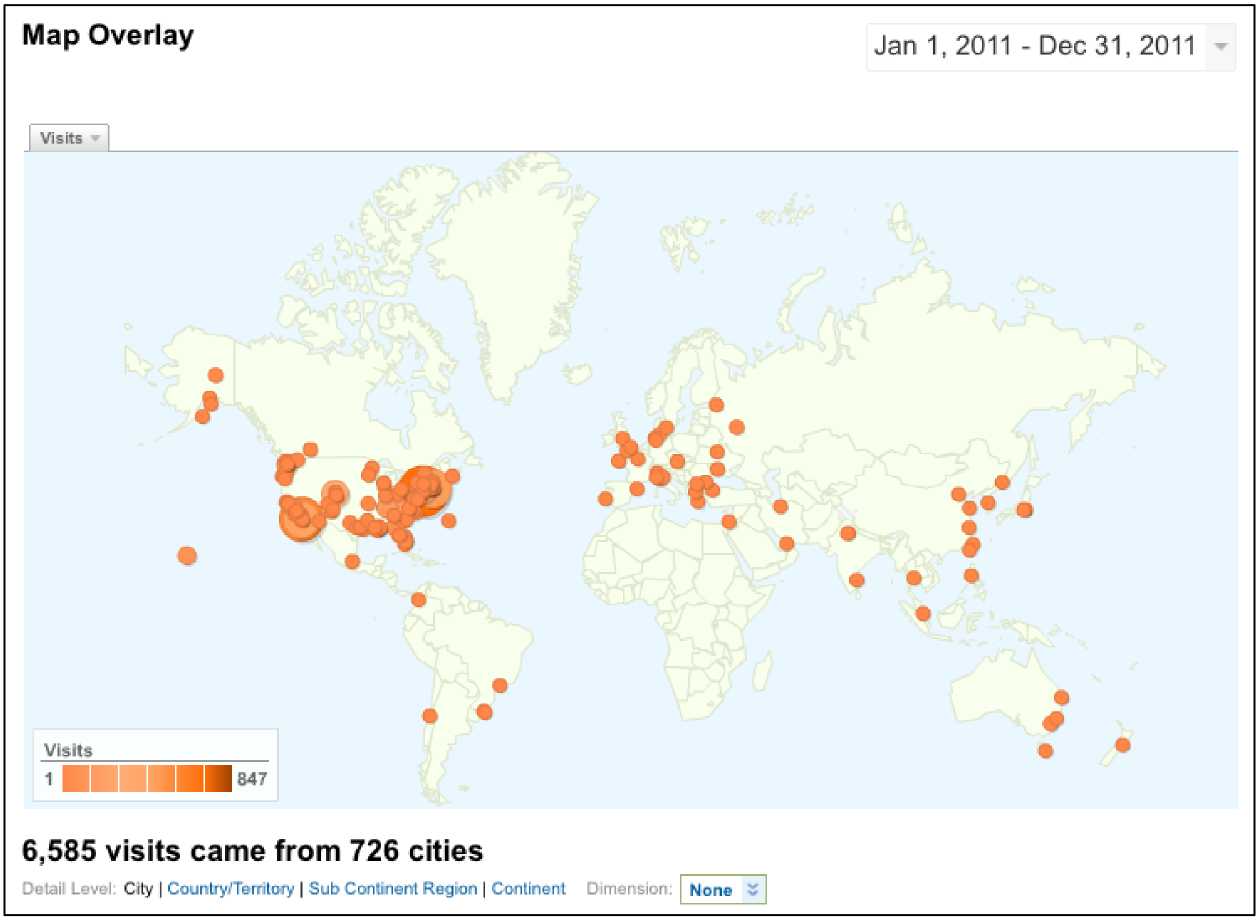


***Figure 5: Website visits as a proxy for users of the R2R Data Repository.** These data represent calendar year 2011 and were collected using Google Analytics (image courtesy Lamont Doherty Earth Observatory).*

Example 3: NEESHub Users

In contrast to the EarthScope PBO and ARF examples, NEES users who wish to download data and interact with resources available on NEESHub are required to complete a simple registration form.[36] Access is free and users create a login and password for future use. The registration form asks users to

[36] The registration form is available at https://nees.org/register.

input their name, NEES Affiliation (i.e., one of the 14 NEES sites), and basic demographic information such as citizenship, gender, and ethnicity. IT administrators at NEESComm (the headquarters for NEES operations) use user registration information cross-referenced with IP address information to better understand the activities of users on the NEESHub.[37] For example, their analyses show increases in user contributions to the NEESHub Project Warehouse (Figure 6); the formation of new collaborative groups; and the extent of WebEx technology use over time.[38]

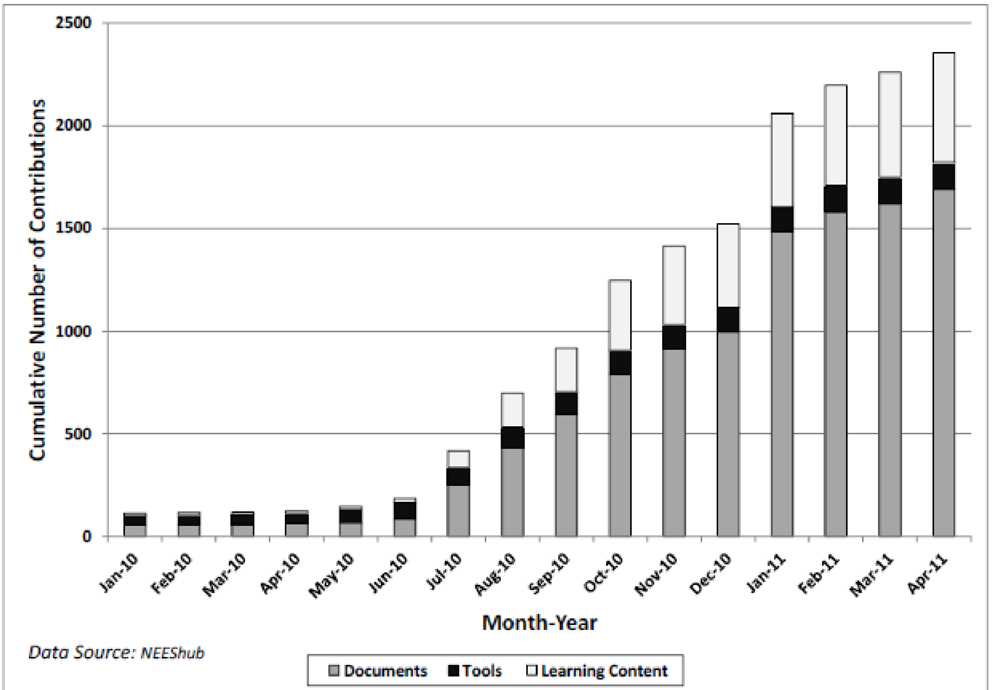


Figure 6: Contributions to NEESHub have Increased Over Time. *On the NEESHub, the total number of contributions (in the form of documents, tools, and learning content) has increased over time, showing the community's adoption of NEES CI resources for their research in earthquake engineering. NEESComm uses the number of contributions to the NEESHub as a method for measuring use of content available on NEESHub (from Hacker and Magana, 2011).*

The data also show *persistent* use, where users continuously return to NEESHub and ultimately adopt it as "an integral part of their ongoing work" (Hacker and Magana, 2011). While the NEES IT administrators are working towards analyzing the demographic information collected through NEESHub user registrations, it is clear that use of a registered username and login enables managers to more powerfully and completely analyze facility use than what would be possible by looking only at IP addresses.

Finally, as some examples of how CI-enabled facilities that are currently under construction *intend* to track users, NEON will adopt an open data policy and will ask users to complete a simple registration form (name, contact information) and agree to a terms of use policy to gain access to NEON data. For

[37] The Hacker and Magana 2011 paper is an outstanding resource for analyzing, interpreting, and applying patterns of use for a facility's CI. The authors clearly and systematically define four stages of how users engage, discover, explore, and ultimately adopt the regular use of CI resources for their research and the paper should be used as a model for other CI-enabled facilities.
[38] Hacker and Magana, 2011.

ALMA, data that is not within its proprietary period (12 months after collection) will be publicly accessible through its online Science Portal without user registration.[39]

3.2 Users On Site

Some facilities track their users by documenting users who *physically* visits the facility to conduct their research and/or interact with the facility. When users also include participants in courses (such as short courses on data processing, etc.), these users are also usually counted. These numbers are typically reported in different sections of quarterly or annual reports to NSF.

Example 1: ARF Cruise Reports

For the ARF, each operator completes a cruise report at the end of each expedition to capture information on ship utilization. This form is requested by UNOLS (not NSF) and is not submitted to NSF: UNOLS maintains these records and uses them to provide information to NSF upon request. As part of this report, each participant is recorded in a spreadsheet containing information such as name, affiliation, and function onboard (e.g., chief scientist, graduate student, technician, educator, observers (observers can be individuals such as science writers, web designers, and marine mammal observers) and foreign observers are individuals onboard to satisfy research clearances required of a host country)). In addition, information on the primary disciplines of research (e.g., biological oceanography, marine geology, etc.) is also recorded. The primary caveat to this dataset is that it relies entirely on the operator to complete the survey – for many expeditions, data are incomplete or not available, making year-to-year analyses difficult. In Figure 7, the total number of users reported is normalized to the number of sea days reported to determine a relative percent of the different user types (as defined by UNOLS, not this study) over a decade.

Example 2: NEES PIs

For the NEES sites, the primary on-site users are the principal investigators (PIs) and their graduate students and/or post-docs. Because experiments can run on the order of three months in duration, the number of projects and on-site users per year is typically low. For example, the NEES-MAST Lab at the University of Minnesota has had a total of nine projects completed with three more in progress since it began operations in April 2005 (where "projects" are differentiated by NSF award numbers and all research at the MAST lab has been NSF-supported, meaning that no outside projects have been scheduled).

[39] See the ALMA website at https://almascience.nrao.edu/alma-data

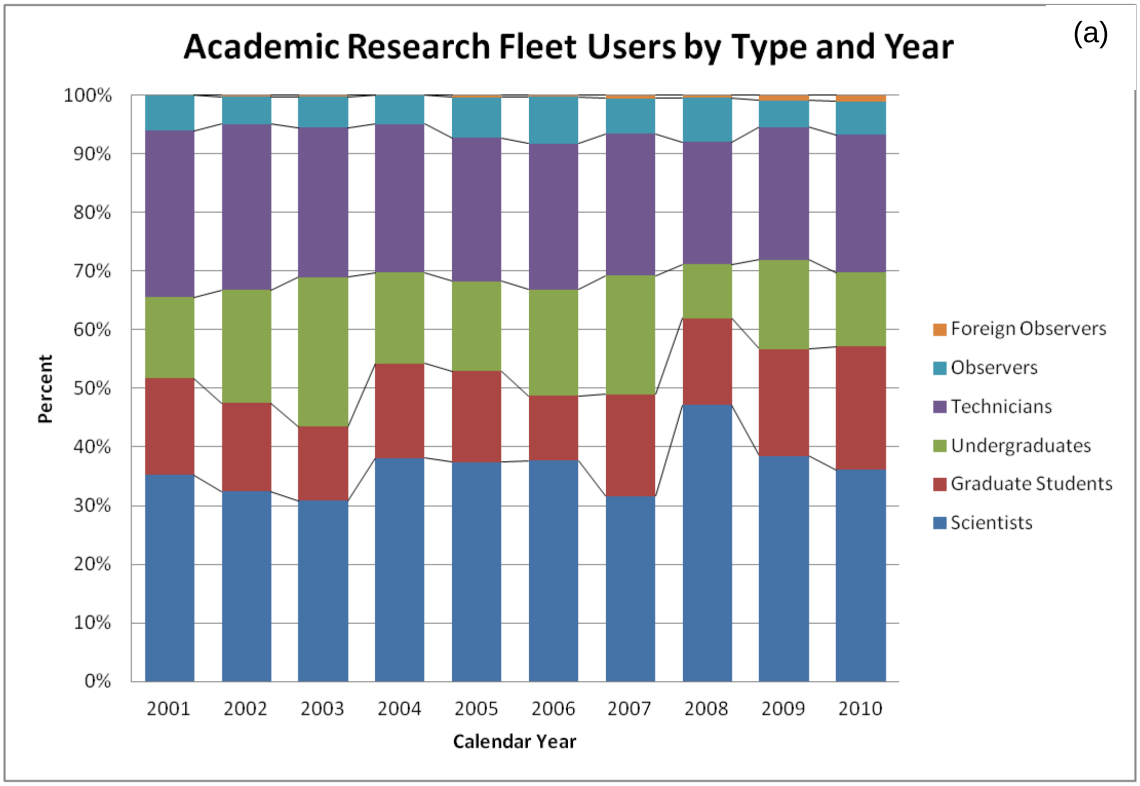
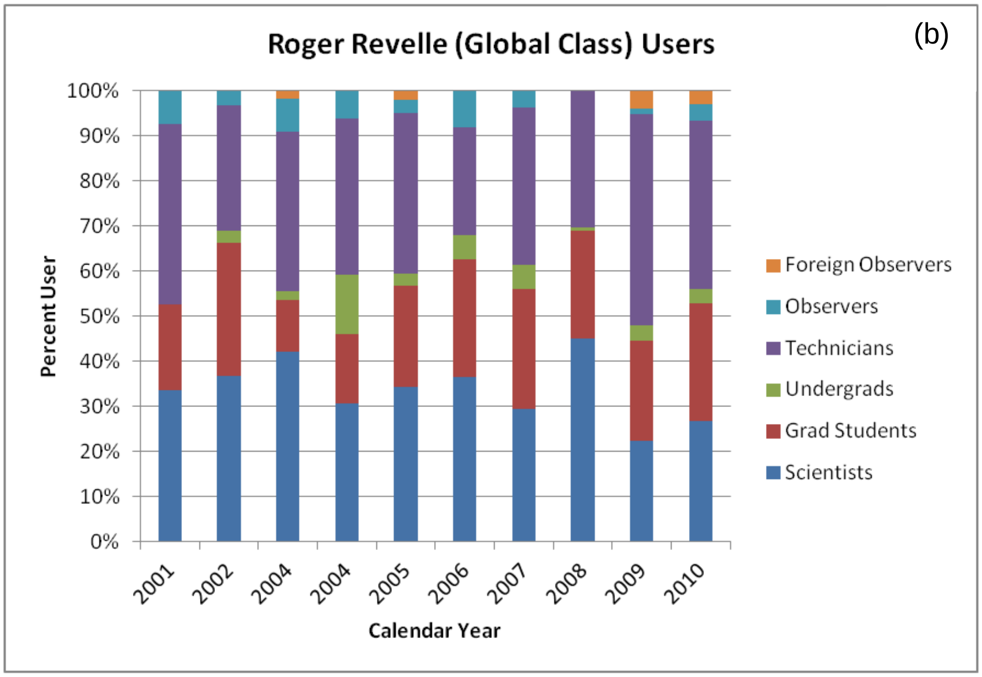
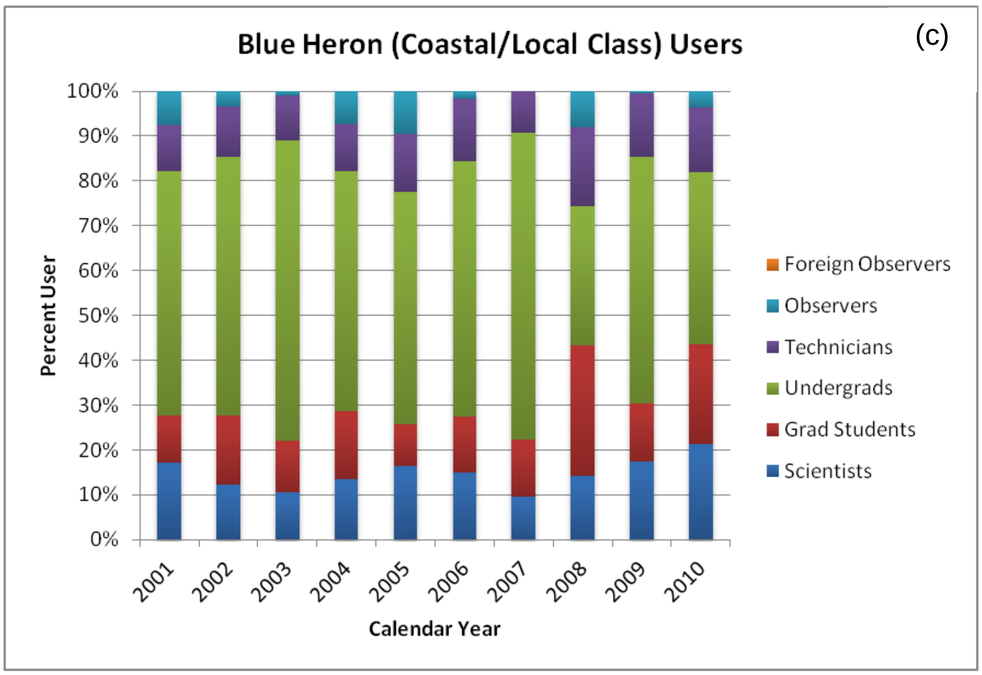


Figure 7: User types over time for the Academic Research Fleet (ARF). *These graphs were produced by the author using data provided by UNOLS, from the Post-Cruise Questionnaire completed by the operators. (a) For each year, the sum and percent of each user type per year was calculated. Not every operator submits a questionnaire for every cruise: this graph was calculated using available data and represents all ships that completed the questionnaire (which includes vessels in the global, intermediate, regional and coastal/local classes, described in Appendix B). Although the total percent of technicians (~20%) appears high, the "technician" category can include* both *ship and research technicians and this may be reported differently for each cruise. While there have been slight increases or decreases in the user types over the years, the profile has largely remained the same for the ships reported*.





Differences in vessel use are revealed by examining data by ship class (e.g., global vs. local).

(b) For global class vessels (those that are >235 feet in length), the distribution of user types is more even.

(c) In contrast, smaller vessels in the coastal/local class (vessels that range from 66-125 feet in length) are more commonly used for undergraduate training, as shown by the relatively high percentage of undergraduates on board each year.

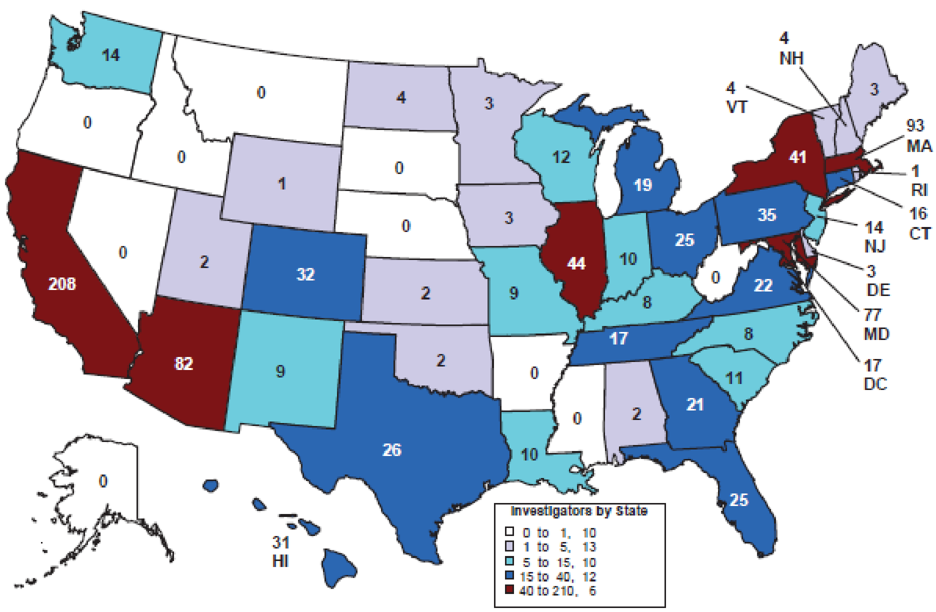
Example 3: NOAO/KPNO Observers

NOAO/KPNO investigators are awarded telescope time through a merit review of submitted proposals. Time is awarded by the NOAO Telescope Allocation Committee (TAC), which meets every six months to review proposals and award time on a semester basis. Applications for telescope time at NOAO facilities are welcome from all astronomers and students, regardless of funding source or nationality (applicants from non-US institutions must provide a justification for using US facilities).[40] Demographic information (institution, role of observer (e.g., investigator vs. graduate student)) about each investigator who is awarded time for on-site observing is recorded in the proposal to the TAC. These numbers appear in the NOAO annual reports to NSF and an example from the NOAO annual report is in Figure 8.

Demographics for the 422 unique observing programs and their investigators are provided below.

Annual Summary Data for Semesters 2011A/B Observing Programs (Excludes NOAO staff except for unique observing programs)		
Description	**US**	**Foreign**
Unique NOAO TAC observing programs scheduled on NOAO telescopes (includes programs under TSIP/FIP on private telescopes)	370	52
Investigators (PIs + Co-Is) associated with approved observing programs	965	422
PhD thesis observers	80	24
Non-thesis graduate students	106	31
Discrete institutions represented	180	167
US states represented (including District of Columbia)	41	NA
Foreign countries represented	NA	31

Breakdown of Investigators from US Institutions for Approved 2011A/B Observing Programs
(Excludes NOAO Staff)



[1] Previous reports incorrectly included data on Chilean and partner institution programs that did not go through the NOAO TAC.

Figure 8: Demographics of NOAO Users. *The NOAO annual reports to NSF indicate demographic information on NOAO users at both KPNO and CTIO (NOAO FY2011 Annual Report).*

[40] See http://www.noao.edu/noaoprop/help/policies.html for additional information.

3.3 Course Participants

All of the case study facilities offer different kinds of short courses and workshops to their users. For scientists and graduate students, these courses are designed to introduce new users to the facility (see Section 5.1) and to train current users in specific tools and resources available from the facility. For example, UNOLS conducts chief scientist training cruises for early career scientists interested in learning how to use ARF vessels (described in Section 5.1); EarthScope offers courses in data processing and equipment use; and NEES facilitates a variety of webinars and short courses for both scientists and practitioners through its NEESAcademy.[41] While NOAO/KPNO does not offer regular short courses, the Virtual Astronomy Observatory (VAO, which includes data from KPNO) has recently coordinated "Virtual Observatory Community Days" at several locations to introduce research astronomers to the VAO and its tools.[42]

Table 5: Example Facility Short Course/Workshop Participants

Facility	Example Short Course/Workshop	# Participants	Report Location
ARF	2011 UNOLS Chief Scientist Training Cruises	28	2011 UNOLS Annual Meeting Presentation,[43] Final Award Report to NSF (Report 1125396)
EarthScope	2006 GPS Surveying and Processing Course	20	2006 EarthScope Annual Review[44]
NEES	NEESHub Boot Camp Webinar (February, 2011)	>75	2011 Q2 NEES Quarterly Report[45]
NOAO/KPNO	VO Day in Tucson, AZ (March 13, 2012)	~50	VAO blog[46]

While short course, workshop, and webinar attendees constitute users, their numbers are reported in different locations, typically in quarterly or annual reports (Table 5). Some participants may also be investigator user types (e.g., downloads data, etc.) while others may be new to the facility. Most of these programs require some form of registration and this information could be cross-referenced with other user data (e.g., login information to an online data archive) to determine whether or not the user is the same individual who is accessing the facility through different means. However, this type of detailed user analysis has not been completed at the case study facilities, to the author's knowledge.

3.4 Event Visitors

Similar to workshop participants, the number of individuals who participate in facility public events, such as open houses and tours, is recorded and reported in different locations. For example, some of the vessels in the ARF occasionally hold open houses and invite the public to tour the ships; EarthScope provides tours of its instruments and sponsors booths for public events like the USA Science and Engineering Festival (many other facilities, including NEES, participate in this event); NEES sites give

[41] See http://nees.org/education/for-professionals for additional information.
[42] See http://www.usvao.org/2012/02/22/vo-community-day-in-tucson-az/
[43] A presentation reporting on the 2011 training cruises from the 2011 UNOLS annual meeting is available online at http://www.unols.org/meetings/2011/201110anu/201110anuap29.pdf and provides details on participant demographics, training agenda, and impressions from the participants provided through a post-cruise questionnaire. Because this effort was funded by an NSF award to a PI (i.e., outside the cooperative agreement for the research vessel on which the cruise took place). The written report on the outcomes of this award was provided by the PI (NSF Award 1125396).
[44] http://www.earthscope.org/es_doc/reports/Year3AnnualReport.pdf
[45] See https://nees.org/neeshubbootcamp for a description and the report at http://nees.org/site/resources/pdfs/2011%202nd%20Quarter%20report.pdf
[46] See http://astrocompute.wordpress.com/2012/03/26/the-vao-community-day-in-tucson/.

tours to school groups and practitioners, and KPNO has a vibrant public outreach program with multiple activities for a wide variety of groups ranging from local school children and the public to programs specifically designed for the local tribes from which KPNO leases its land. Some of these events are held on a regular basis, while others are only on occasion.

While many interviewees disagreed on whether or not these participants constitute "users" of the facility, they interact with facility resources and staff to learn about the facility's science. Public events are also an important component of the pipeline for new users (see Section 5.1).[47] Many of these efforts form a critical component of the facility's fulfillment of NSF's broader impact requirements and is an important part of community building in the facility's geographic area. Quantitative tracking of these types of participants is often done by approximation, especially for large events, and is typically reported in the education and outreach sections of quarterly and/or annual reports (Table 6).

Table 6: Approximate Numbers of Participants at Example Facility Events

Facility	Example Event	# Participants	Report Location
ARF	2005 Open House of R/V *Thomas G. Thompson* at Univ. of Washington	>800	Not in written report.
EarthScope	2011 Arizona State Univ. (ASU) Homecoming Block Party and Earth and Space Exploration Day	>3,000	Spring 2012 *inSights* community newsletter[48]
NEES	2010 USA Science & Engineering Festival NEES booth	~3,000	2010 NEES Annual Report (vol. 1)[49]
NOAO/KPNO	Nightly Observing Program	6,593 in 2010-2011 (46/night)	Kitt Peak Visitor Center Metrics 2010-2011 Report

These events clearly draw hundreds of users, typically of the educator/public/citizen scientist type, even if it is only for temporary interaction with the facility. Because each facility determines its own path and strategy for education and public outreach programming, quantitative user tracking for these types of events is likely only useful for the facility itself (for example, to examine trends in participation over time or different venues or seasons) and cannot be easily compared across facilities.

3.5 Refereed Publications

Across the case study facilities, EarthScope, NEES, and NOAO/KPNO track the total number of refereed publications that use data and/or results from their facility. ARF does not. Most facilities rely on two resources for tracking publications: 1) correspondence from members of the community; and 2) conducting keyword searches in bibliographic databases such as the ISI Web of Science, Google Scholar, and Science Direct. The reliability of community-provided information on new publications is variable. For example, EarthScope asks its community to send publications and supplementary material to the EarthScope National Office and complements these submissions with web searches as a proxy for the total number of publications. In contrast, NOAO/KPNO has a library staff that compiles publications from the community and NOAO staff members. With any scientific publication, the extent to which *all* of

[47] For example, read the "Community Spotlight" section on p. 7 of the Fall, 2011 IODP community newsletter, *Core Discoveries*. This article describes a scientist who toured the IODP research vessel during an open house and has since sailed on five IODP expeditions. Available online at: http://www.oceanleadership.org/wp-content/uploads/2009/03/CoreDiscoveries_Fall2011_WebQuality.pdf
[48] Available online at http://insights.asu.edu/outreach.html
[49] Available online at http://nees.org/site/resources/pdfs/annual_report2010Vol1.pdf

the authors are users of the facility is likely quite variable,[50] but because publications indicate contributions to scientific knowledge, they are a vital part of tracking and assessing facility use and value.

3.6 Tracking Methods Used by Other Agencies

DOE Basic Energy Sciences (BES)

DOE's Office of Science/Basic Energy Sciences has a rich data archive of quantitative user data and statistics. All BES facility users complete a BES Facilities Questionnaire after working at the facility. The explicit definition of a user (see Chapter 2) helps to determine who completes the questionnaire. Across light source, neutron source and electron beam facilities, all users answer the same questions, allowing DOE to collect consistent, statistically relevant data of facility use over time. The questionnaire has been in use since 1990 and user statistics are reported annually. An example is shown in Table 7.[51]

[50] Zhang et al. (2011) present a very detailed study of publications and data use from a variety of astronomy data sources such as the Sloan Digital Sky Survey (SDSS).

[51] Source: http://science.energy.gov/~/media/bes/suf/pdf/BES_Facilities_Number_of_Users.pdf.

Table 7: Number of Users Reported by DOE BES User Facilities

Number of Users Reported by BES User Facilities

Number of Users*

FY 2000	FY 2001	FY 2002	FY 2003	FY 2004	FY 2005	FY 2006	FY 2007	FY 2008	FY 2009	FY 2010	FY 2011	
												X-ray Light Sources
2,551	2,523	2,413	2,206	2,299	2,256	2,105	2,219	2,128	2,214	2,229	2,313	• National Synchrotron Light Source
895	907	1,023	867	741	1,007	1,124	1,151	1,147	1,361	1,436	1,515	• Stanford Synchrotron Radiation Lightsource
1,036	1,163	1,385	1,662	1,898	2,003	2,158	1,748	1,938	1,918	2,032	1,931	• Advanced Light Source
1,527	1,989	2,299	2,767	2,773	3,215	3,274	3,420	3,279	3,537	3,796	3,986	• Advanced Photon Source
										359	516	• Linac Coherent Light Source
												Neutron Scattering Facilities
-	-	-	-	-	-	-	24	165	307	430	890	• Spallation Neutron Source
153	-	22	51	48	96	42	72	258	358	375	477	• High Flux Isotope Reactor**
25	122	164	269	339	221	297	272	261	416	325	308	• Manuel Lujan Jr. Neutron Scattering Center
												Nanoscale Science Research Centers
-	-	-	-	-	-	139	309	404	317	360	374	• Center for Nanophase Materials Sciences
-	-	-	-	-	-	-	164	303	209	274	327	• Molecular Foundry
-	-	-	-	-	-	-	189	272	354	358	348	• Center for Integrated Nanotechnologies
-	-	-	-	-	-	-	112	196	305	377	368	• Center for Nanoscale Materials***
-	-	-	-	-	-	-	-	106	213	281	363	• Center for Functional Nanomaterials
												Electron-beam Microcharacterization Centers
83	88	103	95	128	154	140	199	153	155	190	220	• Electron Microscopy Center for Materials Research
201	212	232	253	241	232	205	183	152	149	164	188	• National Center for Electron Microscopy
99	97	111	112	109	150	132	159	144	161	165	210	• Shared Research Equipment Program

*** Users are researchers who propose and conduct peer-reviewed experiments at a scientific facility.**

• The primary type of user is a Badged User, i.e., a researcher who conducts experiments within the facility.

• There are two other types of users who conduct experiments: (1) Remote User—a researcher who has been granted authority to remotely produce data (this excludes persons who can "look at data"); and (2) Off-Site User—a researcher to whom the facility provides custom-manufactured materials, tools, or devices that the facility has unique or unusual capabilities to fabricate (this applies only to such activities at Nanoscale Science Research Centers). For both types of these users, only one user is to be counted per proposal regardless of the number of co-investigators, and only if no individual is counted in any of the other user categories under the same proposal.

• For annual totals, an individual is counted as 1 user at a particular facility no matter how often or how long the researcher conducts experiments at the facility during the fiscal year. A Badged User cannot also be counted as another type of user. Users must submit a successful, peer-reviewed research proposal and conduct experiments, as described above. Therefore, users do not include individuals who only send in samples to be analyzed, even if such activities are part of a peer-reviewed experiment. Users do not include individuals who pay to have specialty services performed or visit the facility for tours or educational purposes. Users also do not include researchers who collaborate on the proposal or subsequent research papers but do not conduct experiments at the facility.

** The High Flux Isotope Reactor (HFIR) was down for maintenance, safety standowns, and upgrades for significant periods during FY01–FY07. HFIR's users include researchers who perform neutron scattering (figures shown above). HFIR also delivers services such as neutron activation analyses and materials irradiation.

***CNM user counts revised in January 2012 from 470 to 377 for FY10 and from 528 to 368 for FY11.

However, while BES collects user data across its facilities, the same data are not collected for DOE's facilities outside BES. For example, the Atmospheric Radiation Facility (ARM), which is in DOE's Biological and Environmental Research (BER) division, makes much of its climate data freely online to registered users. In turn, BER examines the database of registered users to assess how the online facility resources and data are being used. An example of the ARM user statistics is available at http://www.arm.gov/about/stats/hxvisitors.

Space Telescope Science Institute (STScI) - Hubble Space Telescope (HST)

STScI uses a variety of methods to track users of the HST, including publications, an award database, and online data access. The primary method that STScI uses to track the scientist and graduate student users of the HST is through refereed publications. STScI has a dedicated staff of three librarians who are committed to cataloging publications. Since about 1990, when HST was launched, they have identified and tracked publications that *directly* use HST data. These publications are differentiated from those that only *mention* HST data and the publications are reported differently in their statistics. For cataloging purposes, publications are tagged with information such as the instruments that are used, the number of unique authors, and HST programs (such as large surveys) associated with the publication. The library staff has recently started tracking dissertations that are produced using HST data. Because some STScI staff are directly involved in research using the HST, their publications are tracked differently, and are used for staff evaluations and tenure considerations. Through these efforts, STScI is able to identify trends in the types of data used over time (see Section 5.3).

In addition to publications, STScI tracks users through the total number of investigators who are awarded time to use the HST (Figure 9). Through these efforts, STScI managers can examine similarities and

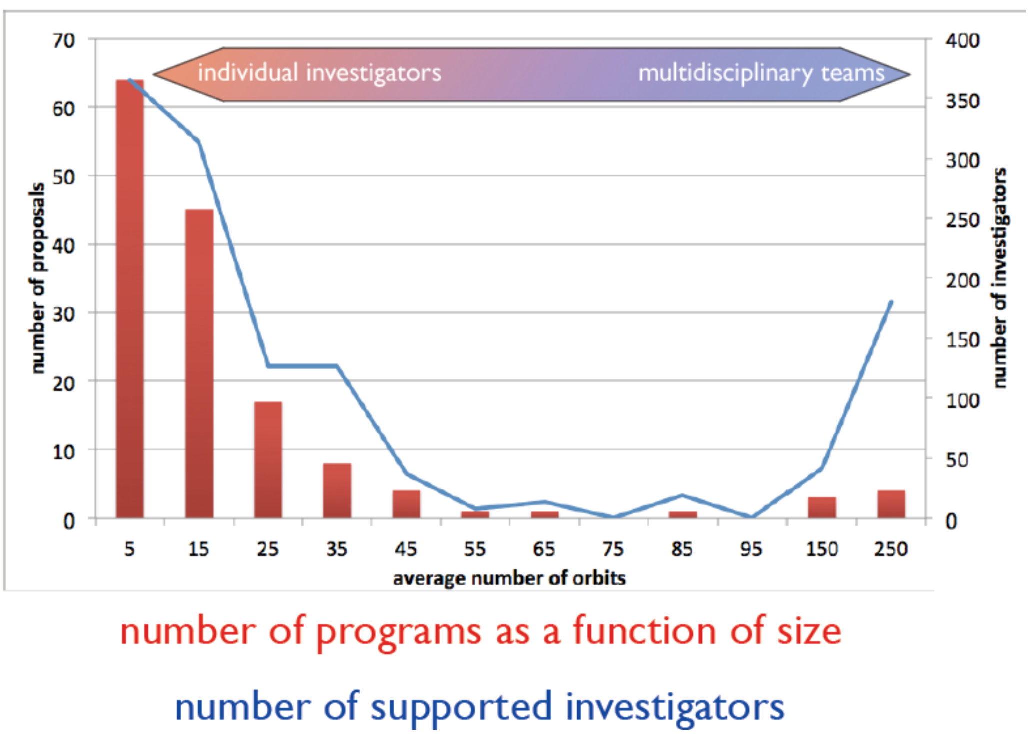


Figure 9: Hubble Space Telescope (HST) program size by number of orbits and number of users. *This image shows the total number of HST orbits as a function of both the number of proposals and the number of investigators using the HST. STScI use this type of data to analyze the relative balance*

between "small" PI-driven projects and large, multi-disciplinary team projects that require a large number of orbits spread over several years.[52] (courtesy STScI)

differences in the number of investigators who are using Hubble time (allocated in number of "orbits"), and how this relates to the types of observing being done (e.g., if the projects are mostly driven by individual investigators using tens of orbits or large, multi-disciplinary teams requiring hundreds of orbits). Finally, HST users are approximated using IP addresses of computers downloading HST data. Similar to EarthScope PBO user tracking, this information can show how many IP addresses are accessing HST data and the geographic location of these computers.

3.7 Discussion

This chapter tackles the question "who is using NSF's large facilities and how are they using these facilities?" on a quantitative level by showing data from a variety of user tracking methods employed by the case study facilities. This chapter has shown that first and foremost, at NSF, each facility tracks its users differently and this information is reported in different formats in a variety of locations.

Each method of user tracking discussed above- IP addresses, data downloads, users on site, course participants, and event visitors - has its caveats. The number of users tracked by IP addresses is at best only an approximation and is likely *underestimating* the actual number of users. For CI-enabled facilities (and facilities with online data archives), the combination of a registered user login and an IP address is a more powerful tool than just an IP address for analyzing user interactions with the facility. Facilities that are accessed physically by their users typically track the number of users on site. This is done using a variety of methods and is not always part of NSF's reporting requirements : these numbers are therefore not easily attained. Facilities that track users by recording the number of participants in short courses/workshops/webinars and in facility-coordinated public events report these values in different locations. Because the workshops and public events are unique to the facilities, and often vary from year to year, this information is the most useful to facilities themselves, and comparison across facilities is unlikely to be informative. Tracking the authors of refereed publications is at best only a proxy for the total number of users of a facility.

User tracking data provides the most *quantitative* information available on the use of a facility and in some cases can be a good indicator of the "health" or vitality of a facility, depending on what is tracked. However, it does not necessarily capture duration of use, scientific output, or scientific value of the research and/or education efforts underway at the facility and any quantitative information on use should always be considered in context. In addition, because of the differences in terminology and tracking methods across facilities, it is not practical to quantitatively compare use across NSF's large facilities.

While user tracking provides valuable insight into the *numbers* of users interacting with a facility, perspectives on the quality of a user's experience with a facility and how the facility is valued by the community are best derived from user feedback to facility management. The next chapter explores the research questions from the perspective of user management and the user's interaction with the facility and the scientific community at large.

[52] This slide was presented at the "Implementing Portals to the Universe: Best Practices for NASA Science Operation Centers" workshop held at STScI on April 25, 2012. The workshop was designed to respond to a 2007 NRC Report on NASA Science Centers (Portals to the Universe: The NASA Astronomy Science Centers." Workshop slides and the final workshop report are available at http://www.stsci.edu/~inr/portals.html.

4. Investigating Users and Facility Use Through User Feedback

User feedback to facility management can provide insights into how facilities are accessed and used. Across the case study facilities, there are multiple mechanisms in place for users to provide feedback to the facility on facility performance and user needs. Agency-wide, NSF gleans valuable information about its facilities and users through proposals (both those that are awarded and declined funding), the development of science plans and/or strategic plans that are formulated during the planning and operational stages of the facility, committee of visitors[53] reports, and advisory committees. With respect to user feedback directly to facilities/awardee institutions, there are many mechanisms for the users at large to provide feedback to the facilities. These include a formal user committee (or committees); informal feedback; online outlets; different types of evaluations or user surveys; and periodic community-wide surveys (Table 8). Each of these methods is discussed in this chapter.

Table 8: Mechanisms for User Feedback

Facility	User Committee(s)	Informal Feedback	Online Feedback	Evaluations	Community Surveys
ARF	✓	✓	✓	✓	✓
EarthScope	✓	✓	✓		
NEES	✓	✓	✓	✓	
NOAO/KPNO	✓	✓	✓	✓	✓

4.1 User Committees

Most facilities across NSF have some form of a user committee, which liaises between facility management and the science user community at large. In this study, a user committee is defined by three characteristics: 1) the committee is composed of community members (typically researchers) who use or have used the facility; 2) the primary purpose of the committee is to represent and communicate the needs of the users to facility management. It is more focused on practical matters of facility use and operations than it is on providing advice, strategic planning and/or facility oversight; 3) the committee meets on a regular basis.

At NSF's large facilities, user committees typically have a chair or two co-chairs and members typically serve a fixed term of a few years. Some facilities have multiple user committees to address the needs of different functions of the facility and/or different sub-disciplines of users. Table 9 provides a complete listing of user committees for each of the case study facilities.

As a comparison to NSF, the Department of Energy (DOE) Office of Science describes its facility "Users' Executive Committees" (UEC) as groups that "foster information exchange and communication on issues that are of interest to the user community."[54] Each DOE facility is required to have a user committee of this kind and each facility has only one UEC.

The HST, which is operated jointly by STScI and Goddard Space Flight Center (GSFC) has one users committee: the Space Telescope Users Committee (STUC).[55] The STUC provides user advice to the observatory as a whole and has a membership of approximately twelve scientists, chosen jointly by the STScI Director and HST Project Scientist. Members serve three-year terms and include PIs of HST observing programs and members of the HST Instrument Definition teams. The scope of the Users

[53] See http://www.nsf.gov/od/oia/activities/cov/ for more information.
[54] See http://science.energy.gov/user-facilities/frequently-asked-questions/
[55] Additional information, including archives of past meetings since HST was launched in 1990, is available at http://www.stsci.edu/institute/stuc.

Committee is quite broad: the STUC can set its own agenda and if necessary can establish an ad-hoc subcommittee structure to perform in-depth studies of relevant subjects.

Table 9: Case Study Facility User Committees

Facility	# User Committee(s)	Committees
Academic Research Fleet	6[a]	Arctic Icebreaker Coordinating Committee, Deep Submergence Science Committee, Fleet Improvement Committee, Scientific Committee for Oceanographic Aircraft Research, Marcus Langseth Science Oversight Committee, and the Ocean Observing Science Committee
EarthScope	7[b]	Plate Boundary Observatory (PBO) Advisory Committee; US Array Advisory Committee; SAFOD Advisory Committee, PBO Data Products Working Group, the Electromagnetic Working Group; Transportable Array Working Group, and SAFOD Core/Sample Working Group
NEES	3[c]	Users Forum, Data and Curation Subcommittee, Equipment Site Forum
NOAO	1[d]	The NOAO Users Committee

[a] see http://www.unols.org/committees/index.html;[b] see http://www.iris.edu/hq/about_iris/governance and http://www.unavco.org/community/governance/committees/committees.html; [c] see https://nees.org/about/contact/neescommteam;[d] see http://www.noao.edu/dir/usercom/

4.2 Informal Feedback

Most interviewees cited informal feedback as an important mechanism for learning about needs and concerns of the user community. The primary path for informal feedback is through scientific meetings and conferences, such as the American Geophysical Union Fall Conference, the biennial EarthScope National Meeting, IRIS and UNAVCO Workshops,[56] the annual NEES National Meeting, Earthquake Engineering Research Institute (EERI) annual meeting, the UNOLS Annual Meeting, and the annual meeting of the American Astronomical Society. At these conferences, scientific sessions, town hall meetings, and facility exhibit booths all provide venues for conversations between facility management, funding agency representatives, and users.

For some facilities, informal feedback takes place via direct phone calls. For example, EarthScope PBO personnel at UNAVCO described instances of users (in this case, commercial surveyors using GPS data) calling UNAVCO with questions on how to use online data. In other cases, UNAVCO sometimes receives calls from public citizens who noticed damage to a GPS station and called to report their observations.

Across the case study facilities, interviewees agreed that informal feedback is important to facility operations. Some commented that informal feedback can provide added value in creating personal connections between facility support or administrative staff and the users.

4.3 Online Feedback

All of the case study facilities have different forms of online mechanisms for users to contact the facility with comments and/or questions. For the ARF, the "UNOLS Office Feedback Form" (http://www.unols.org/forms/_feedback.asp) is available to users who wish to provide comments on "regarding the UNOLS website, meeting announcements and support, travel guidelines and any other

[56] These alternate: EarthScope National Meeting is in odd-numbered years, and the IRIS Workshop and UNAVCO Science Workshop are held in even-numbered years.

items of concern." For EarthScope, there are multiple online options for users to request support, ask questions and/or provide feedback on data access to the facility operators (e.g., http://achaia.unavco.org/public/newproject/supportform.aspx and by clicking on "User Feedback" in UNAVCO's Data Archive Interface). For NEES users, the NEESHub offers both support ticket submissions and an online user forum for users to ask the community for assistance. NOAO invites users to email them at the Virtual Observatory for help in accessing data. They also include an online tutorial and guide to new users.

4.4 User Satisfaction Surveys

Similar to customer surveys issued by businesses, three out of four of the case study facilities employ some form of user evaluations or user surveys to assess the user experience with the facility. The survey questions are typically written by facility managers; although some facilities contract a survey professional to write, collect, and analyze survey data. The case study facilities that use these types of evaluations agreed these surveys are an important component of user feedback. However, across the facilities, every form is different.

For the ARF, UNOLS has an online "Post Cruise Assessment Report" form (http://strs.unols.org/Public/diu_pre_pcar.aspx). One UNOLS administrator described the form as something "that has been in place in place for many years" and remarked that, "although the form is not mandatory, chief scientists are strongly encouraged to submit a form after every cruise. Captains and marine technicians are also asked to complete the form after every cruise." EarthScope has no known user satisfaction surveys, although IRIS and UNAVCO (EarthScope's awardee institutions) may have informal surveys for their campaign instrumentation pools. NEEScomm issues a "NEES User Feedback Survey" to "all PIs with active or recently completed projects that involved work at NEES facilities" and results are collected by a third party (in 2011, the report was compiled by the University of Michigan's School of Information) and submitted to NEEScomm in aggregate form.[57] NOAO/KPNO collects "end of run" evaluations to capture user feedback from their experience observing at KPNO. A blank form is available at http://www.noao.edu/cgi-bin/ore/oreform.pl.

Across the case study facilities, administrators agreed that one of the greatest challenges in administering user satisfaction surveys is "getting people to respond." This topic arose at the 2012 Annual NSF Large Facilities Workshop[58] and while many agreed that surveys could be useful, there were few conclusions on what types of surveys (e.g., paper vs. online), and distribution of surveys (e.g., while user is on site vs. off-site) produce the most beneficial results.

By comparison, DOE includes a user satisfaction survey in their Basic Energy Sciences (BES) Annual Facilities Questionnaire. The same questions are asked of users at all BES facilities; but not across DOE's other user facilities.

4.5 Community Surveys

While user satisfaction surveys form an important tool for assessing an individual user's experience at a facility, they do not provide extensive information about the *community* of users – this is especially true for

[57] See Zimmerman, 2011.
[58] The Large Facilities Workshop is an annual event coordinated by NSF's Large Facilities Office in partnership with a different facility each year. The goal of the workshop is to discuss best practices in facility management and operation: the workshop is open to all and typical attendees include facility administrators, IT specialists, site operations managers, and education specialists. The website for each workshop is hosted by the host facility; the 2012 workshop site is http://meetings.nscl.msu.edu/lfw2012/.

facilities that have multiple types of users accessing multiple facility resources from multiple physical and remote access points. Some facilities lead community-wide surveys to better understand their users, potential users, and their disciplines at large. Historically, this technique has been used by facilities planning future assets (e.g., new telescopes, new data products, etc.) and by facilities facing recompetition or renewal. Results can point to new directions for the facility and new areas of future research for the field. With the advent of online survey technology (e.g., surveymonkey.com, zoomerang.com, etc.), administering online surveys has become easier to conduct. The following discussion presents three examples of community surveys led by NSF large facilities and some of the successes and challenges of interpreting the results.

Example 1: UNOLS Shiptime Demand Survey

In 2010, UNOLS conducted a “shiptime demand” survey of the oceanographic research community to elucidate why there had been a decrease in the number of proposals submitted to federal agencies to use UNOLS vessels (http://www.unols.org/info/vessel_usage_survey.html). The survey was posted on the UNOLS website and distributed through its established networks and email listserves. Recipients of the survey link were encouraged to forward the link to their colleagues to encourage a wide community response. Survey questions ranged from demographics to factors that impeded respondents from submitting proposals for at-sea research. Over 300 people responded to the survey and the majority of the respondents indicated that they intend to submit ship-time proposals in the following year; but the community at large voiced concern about their perceived chances of being funded being less because of proposing to use ships for their research. The survey continues to be a useful tool for connecting with the community and for opening a dialogue between the community and funding agencies.

Example 2: IODP’s Strategies and Priorities Survey

In 2011, the Integrated Ocean Drilling Program (IODP) administered an online survey to gather input from its US user community on priorities for future research in the next program (scheduled to begin in 2013, pending renewal approval). Similar to the UNOLS survey, the IODP survey link was sent via email to established listserves and posted on websites: community members were encouraged to share the link with their colleagues. Questions included research discipline, career stage, and areas of future research interest. A total of 433 individuals responded,[59] showing that 25% of the respondents were early career scientists, and 19% were students. Nearly half (49%) of the respondents were new to the program, showing strong interest for a future drilling program from the community at large. Results of the survey were used to guide an April 2012 community workshop, “Building US Strategies for 2013-2023 Scientific Ocean Drilling.”[60] While facility administrators and community members found the results useful, some expressed concern about the statistical relevance of their survey – did the 433 respondents fairly portray the interests of the IODP user community at large? What percent of the community does 433 represent? (By comparison, over 800 scientists and students participated in their 2009 INVEST workshop[61], and over 1000 email addresses are subscribed to the US IODP listserv database). These questions remain unanswered, but appear common across facilities when these types of surveys are conducted.

---

[59] A report on the survey was submitted to NSF on June 15, 2012. An overview of the survey is provided in IODP’s Spring 2012 community newsletter (p. 6), available at http://www.oceanleadership.org/wp-content/uploads/2011/10/CoreDiscoveries_Spring2012_Final_WebRes.pdf. Selected results were reported at the 2012 Ocean Sciences conference (Meth and Ludwig, 2012).

[60] See http://iodp-usssp.org/workshop/strategies/ for additional information.

[61] The “New Ventures in Exploring Scientific Targets” (INVEST) workshop was held in 2009 and invited the international community to weigh in on the scientific research goals of a new ocean drilling program. INVEST was one of the largest community workshops for the drilling community in the program’s recent history; additional information is available at http://www.marum.de/iodp-invest.html.

Example 3: Ground-based Optical/Infrared (O/IR) Observational Astronomy Survey

In 2011, the US Ground-based Optical/Infrared (O/IR) System Roadmap Committee[62] administered a survey of the astronomy community "to inform our assessment of current ground-based O/IR observing facilities and the community's use of these facilities." Similar to IODP, the survey was also used to collect information on community members' "plans for using existing facilities to pursue science highlighted in the reports of the National Academy decadal surveys" (Jannuzi and Valenti, 2012). The survey was conducted online and asked for information ranging from demographics (e.g., profession, home institution type, discipline) to user needs, the use of existing telescopes, and future needs of the astronomy community. A total of 1,178 responses were received, including 962 based at US (or US-sponsored) institutions. For placing the response rate in context, Jannuzi and Valenti (2012) point to the membership of the American Astronomical Society (6279 members as of 1/31/12). The results of the survey were analyzed by the Roadmap Committee members and reported to NSF.[63]

The results of this survey present a unique picture of user interactions with ground-based astronomy facilities. While the scope of the US Ground-based Optical/Infrared (O/IR) System Roadmap Committee extends to facilities beyond those supported by NSF, the results of the survey hold an important lesson: *users rely on multiple telescopes to complete their work.* This is explicitly illustrated in Figure 10. As of the printing of this report, the results of this survey were still under discussion within the Astronomical Sciences division and Mathematical and Physical Sciences directorate at NSF.

[62] As described in Jannuzi and Valenti (2012), the "US Ground-based O/IR System Roadmap Committee is a standing advisory committee charged by NOAO to assess annually the state of the ground-based O/IR system of observing facilities (i.e., all ground-based O/IR telescopes operated by US institutions, including both federal and non-federal facilities) and to make recommendations regarding which capabilities are needed by the community on near and long term timescales… the Committee has representation from the entire US community that uses the system."

[63] The "Sustaining Progress Toward The Decadal Survey Science Priorities Over the Next Decade" report was submitted to the NSF/AST Portfolio Review Committee by the US Ground-based O/IR System Roadmap Committee in January, 2012. See http://ast.noao.edu/about/committees/system-roadmap

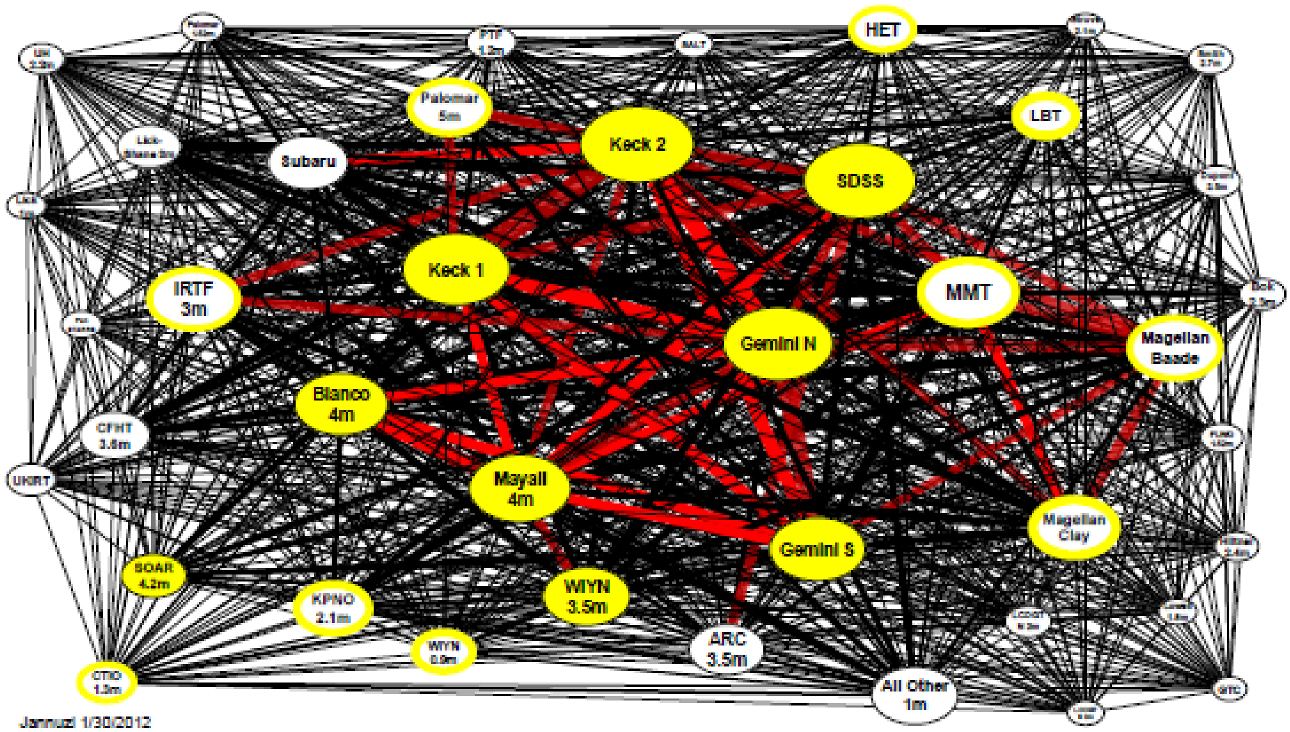


Figure 10: Results from the 2011 Ground-based O/IR Ground-based Astronomy facility survey. *Jannuzi and Valenti (2012) show that users rely on a* system *of telescopes to complete their work. The figure above and text below are directly from the report.*

The US has a diverse and capable set of ground based O/IR observing capabilities enabled through the combined efforts of the NSF, DOE, NASA, and non-Federal observatories and institutions. Shown are all the US telescopes (these facilities are run by US institutions, or have a US partner, i.e. some fraction of the observing time for each of these facilities is allocated by a US run institution) used by more than 3% of US based respondents to our November 2011 survey of the astronomical community. Shown are results from US based respondents. Each telescope is shown as an ellipse whose area is proportional to the fraction of the respondents that reported using that telescope in the last three years. The thickness of lines between the telescope ellipses is proportional to the number of people that used both of the linked telescopes. The largest lines (representing more than 7% of respondents each) are in red to clearly show the strongest connections. While this manner of displaying the survey responses does not adequately show how many people used multiple telescopes, it does graphically demonstrate that the most frequently used telescopes (largest ellipses) are used by astronomers that are also using multiple other facilities. Those using the less frequently used telescopes are also heavy users of the most used facilities. Telescopes that have received, on average, more than $1M per year of support from the NSF for the last 10 years have their ellipses filled in yellow. Facilities that we are aware of having received NSF operations or other support from NSF/AST facilities, TSIP, ATI, MRI, PREST, and ReSTAR at a lower, but still significant, level are shown with a yellow border.

The examples above clearly show that a wealth of information can be gleaned from community surveys, including data that would not be captured in databases such as the NSF award database, or even lists and numbers of facility users. Surveys can illustrate how users interact – and intend to interact[64] – with

[64] In 2009 NEON used a professional survey company called Corona Insights to conduct both an online survey and in-depth interviews of key stakeholders in the ecological sciences community. The goal of the survey was to assess *potential use* of and community awareness about NEON's future resources (NEON had not yet officially entered the MREFC-funded construction stage at this time). Corona's study was guided by three research questions: "What is the awareness of NEON among these key stakeholders? What are their attitudes and perceptions towards NEON? What is their preferred mode of communications? And what is the profile of stakeholders?*"* A total of 2,485 individuals responded to the

facility resources and are an important tool for understanding user needs. However, it is important to recognize that administering surveys by mass email (termed "snowballing" or "chain sampling"[65]) is not a well-controlled study – the statistical relevance of the results can never be fully known and therefore results must be carefully interpreted.

In summary, these examples show that conducting community-wide surveys on facility use (or potential use) can be an effective method for collecting information. However, the results should be used with caution and taken in the context of the survey method and sample size when making policy and/or management decisions.

4.6 Discussion

At NSF's large facilities, users have multiple outlets for providing feedback to the facilities they use. These outlets include user committees, informal feedback to facility management and funding agencies at conferences, online tools, user satisfaction surveys, and community-wide surveys and these are all important tools for understanding users and how users use facilities.

Results from this chapter show that not all facilities employ all of these methods of user feedback and each facility has a different "version" of these feedback tools. With respect to user committees, not all NSF facilities use the term "user committee" for groups that are composed of users and are designed to interface between facility management and the user community. These committees are sometimes, but not always, considered part of the facility governance structure. In terms of informal feedback, including face-to-face interaction, both users and managers felt that this plays an important role in facilitating communication between facility management, funding agencies, and users.

All case study facilities provide online mechanisms for users to contact the facility with comments and/or questions, regardless of whether or not the facility is CI-enabled. Results from this chapter show that user satisfaction surveys are an important tool for communicating user needs and experience to facility

---

online survey, which was distributed to ecologists affiliated with NEON, the Ecological Society of America (ESA), and the American Institute of Biological Sciences (AIBS) (for comparison, ESA has over 10,000 members), representing a 22% response rate. Corona developed the survey in consultation with NEON and administered the survey online. Respondents were controlled by using unique identifiers (usernames and passwords) attached to each respondent so that each person could respond once, and the survey could not be forwarded to other participants. The survey questions included information about the respondent's demographics (race, education, disciplinary expertise) as well as their perceptions of NEON. Results from the survey showed variations in interest and awareness of NEON and how this varied with age and education level: in general, awareness was very high (81% of the respondents were aware of NEON). The study also revealed direct correlations between career stage and interest in using NEON (more senior scientists were in the "uninterested" category while individuals who were interested in NEON but varying awareness of NEON's resources were at earlier stages in their career. Survey results were used to formulate recommendations for improving communication tools and messaging to reach potential users and to build awareness of NEON going forward. NEON is the first known NSF large facility to conduct such a thorough survey of its community and potential users during the planning phase of the facility. The long-term applications of employing such a tool at this stage of the facility development process remain to be seen.

[65] Snowballing is a non-probability sampling technique where existing study subjects recruit future subjects from among their acquaintances (such as forwarding a web survey link via email). One of the advantages of this technique is that the survey may reach people that would have otherwise been unknown to the survey administrator (http://en.wikipedia.org/wiki/Snowball_sampling). Other agencies have used snowballing surveys to conduct user surveys of resources such as satellite imagery and have reported on different techniques that work well and others that may compromise response rate (see http://www.fort.usgs.gov/landsatsurvey/OnlineSurveys.asp)

managers. Across NSF's facilities, user surveys are written, collected, and reported differently. There are no known facility-wide "best practices" in the administration and analysis of user satisfaction surveys.

In terms of placing the facilities in the context of the scientific community at large, NSF's large facilities are uniquely positioned to lead community-wide surveys because of their size, resources, and reach across many disciplines and fields. Community surveys can provide a valuable "census" of scientific communities and survey data can be used to assess the use of a facility, complementary use of other facilities by the same users, and the need for new facility resources in the future. Online survey technology combined with multiple dissemination routes such as email listerves and website posts makes conducting community-wide surveys an easy and cost-efficient process using a snowballing technique. However, in many cases, community surveys are written, distributed, and analyzed by individuals who are not professional survey experts. This may complicate the statistical validity of the results.

Facility managers use results and information provided through user feedback tools to inform decisions regarding resources, new tools, and for communicating with funding agencies. User feedback is a critical tool for managing the relationship between users, facility management, and funding agencies: it can also play an important role in the facility's relationship with the scientific community at large. The next chapter builds upon this by showing changes in facility use over time and how some facility user types, user backgrounds, and user pathways reflect changes in science and technology.

5. Changes in Facility Use Over Time

The development of a large facility, from the planning to the operational phase, can take decades to complete. Many facilities are in operation for decades more after operation begins. Science and technology change over these periods of time, and in some cases, very rapidly. Facilities must be prepared to adapt to these changes where new technologies can be implemented to maximize scientific return while maintaining the research goals of the facility.

When facility stewards, managers, and users were asked about how facilities and their users have changed over time, several themes emerged. These included unanticipated users, an increase in data-driven use, changes in new user skills and backgrounds, and an increase in collaborative or "team-science" projects. These themes are discussed below and each one points to important opportunities and challenges for planning and managing future facilities.

5.1 Incorporating New Users

Across the case study facilities, administrators and managers agreed that it was important to identify, welcome, and accommodate new users to their facility. New users can bring new ideas to the facility, and in some cases, new sources of funding. New users who are early career scientists also are an important part of the future scientific workforce and facility stewards universally agreed in the value of fostering this pipeline.

At NSF, new users discover facilities via both *informal* and *formal* pathways (Fig. 11). All facilities participate in various types of community "outreach" events, regardless of the age of the facility. Some of these events are designed to advertise and promote the facility to potential new users and include booths at academic conferences and lecture series. Other tools are more diffuse and informal: every facility has a website, most have email distribution lists that are open to anyone to join, and some facilities are exploring social media tools as a way to reach both new users and the general public. These are all examples of informal pathways by which users discover the facilities.

Formally, new users find out about facilities through a variety of programs coordinated by the facility. For example, EarthScope offers different of short courses to new users to learn about specific types of field campaign equipment and/or data processing. IODP features a half-day workshop in conjunction with major academic conferences designed to introduce new users to the program.[66] UNOLS teaches a formal "chief scientist training" program every summer as a professional development opportunity for early career scientists interested in using the ARF.[67] Many facilities sponsor undergraduate interns (see Section 2.2), and some of these interns continue on to pursue graduate research using the facility.[68] Some facilities (e.g., NEES, NNIN, IODP) use webinar technology to invite interested participants to

[66] The US IODP began offering its "IODP Primer: An Introduction to Scientific Ocean Drilling" in 2009 and since then, registration for every course has been completely filled. IODP administrators at the US Science Support Program have received positive feedback on these courses, but believe that additional time is needed to determine the long-term impacts of the program. For more information, see http://iodp-usssp.org/workshop/iodp-primer-2011-agu/

[67] See http://csw.unols.org/

[68] The RESESS program, led by UNAVCO in partnership with IRIS, the Significant Opportunities in Atmospheric Research and Science (SOARS), and USGS (http://resess.unavco.org/), has some outstanding success stories of undergraduate interns going on to pursue graduate degrees in the same field.

attend online lectures, courses, or information sessions about opportunities available through their facility.[69]

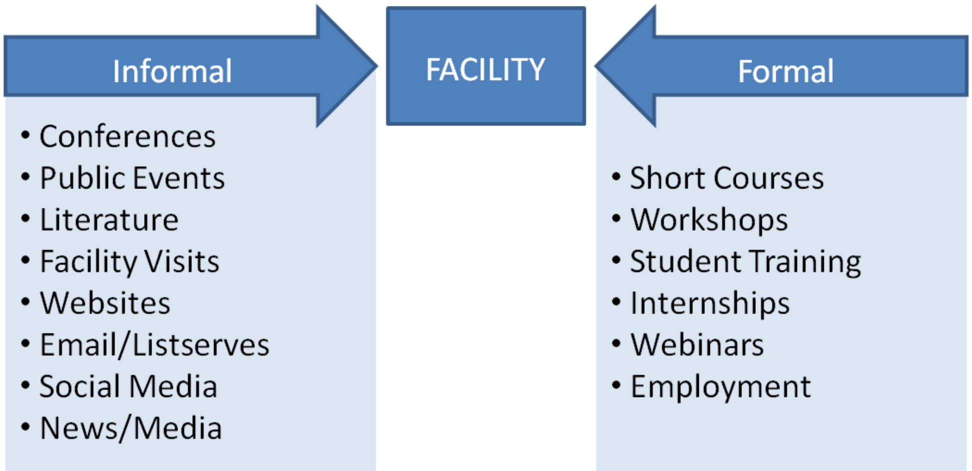


Figure 11: Informal and formal pathways by which new users discover facilities.

The degree to which facilities "actively" seek new users varies, depending on facility type. For example, in the case of the NEES MAST lab, the lab is already over-subscribed through 2014, after which the NEES program will be re-competed. Therefore, the MAST lab did not see the need to invest many resources into bringing in new users. However, when NEES began in 2004, facility personnel from the MAST lab produced brochures, and gave talks at conferences and universities to advertise the availability of the lab and raise awareness about the new facility. Similarly, in the early days of ALMA construction, NRAO held 25-30 "road shows" over the course of 9-12 months to recruit new users and raise awareness about ALMA's resources. These events reached approximately 2000 people.

Facilities trying to identify new scientific users have found success in hosting booths, town halls, and information sessions at conferences outside their immediate field. For example NNIN, which supports multiple sites for nanotechnology, hosted a booth at the 2012 Ocean Sciences meeting to attract new users from the ocean and geological sciences community. The NNIN representative at the booth later followed up directly with booth visitors to invite them to the online NNIN webinars focused on environmental sciences.

5.2 Unanticipated Users

Across the case study facilities, all agreed that the users for whom the facility was intended were using the facility. However, some facilities, like EarthScope, had stories of "unanticipated users," where individuals they never expected to become users of the facility came forth to engage with the facility. This resulted in an expansion of the number of users and user needs, and in some cases, led to new fields of research.

Example: EarthScope's Unanticipated Users

EarthScope was built by and for the solid earth geosciences community. Intended for geologists, seismologists, and geodesists, EarthScope set out to be "an interdisciplinary experiment of

[69] NNIN recently launched a series of webinars on nanotechnology as it relates to other fields. See http://lnf.umich.edu/nnin-at-michigan/index.php/geosciences/workshops-and-webinars-series/

unprecedented resolution that will identify links between the surface geology of North America and the forces at work in the Earth's interior." [70] The 2001 Project Plan for EarthScope describes the facility as:

> "a new Earth science initiative that will dramatically advance our physical understanding of the North American continent by exploring its three-dimensional structure, and changes in structure, through time… by integrating scientific information derived from geology, seismology, geodesy, and remote sensing, EarthScope will yield a comprehensive time-dependent picture of the continent beyond that which any singly discipline can achieve."

Correspondingly, EarthScope's cyberinfrastructure and data products were developed for these disciplines and users. However, after EarthScope became operational in 2004 and made data freely available online, multiple new users discovered and used EarthScope's resources (Figure 12). Hydrologists at the University of Colorado in Boulder began using GPS data to estimate snow depth and soil moisture.[71] Researchers at the Laser Interferometer Gravitational-wave Observatory (LIGO) downloaded EarthScope's seismic data to subtract noise from the Earth's motions in their gravity wave research. Glaciologists have begun exploring EarthScope's seismic data for remote detection calving glaciers. Meteorologists are investigating correlations between strong Midwest storm systems and EarthScope seismic data. NEES researchers have used EarthScope's seismic data to inform their work in earthquake engineering. Beyond academic applications, EarthScope's data managers have received calls and noticed data downloads from public utility companies, who use the GPS data for planning city pipelines and powerlines. Interestingly, many of these unanticipated users are supported by other programs or facilities within NSF (Figure 12).

[70] EarthScope Workshop Report, *Scientific Targets for the World's Largest Observatory Pointed at the Solid Earth*, 2001.
[71] One PI was awarded the prestigious American Geophysical Union Fellowship for her innovative use of GPS technology in this field of research.

USERS	TYPE OF USE
Geodesists*, Seismologists*, Geologists*	**GPS and Seismic data to study structure and evolution of N. American continent**
Educators*/Students	**Real time data and curricula**
Hydrologists*	GPS data to estimate snow depth, soil moisture
Public Utilities companies	GPS data for city pipeline planning
LIGO Researchers*	Seismic data to inform gravity wave research
Glaciologists*	Explore seismic data to detect calving glaciers
NEES Researchers*	Seismic data to inform earthquake engineering experiments

**funded by NSF*

Figure 12: EarthScope's Unanticipated Users. *The EarthScope facility was intended for geologists, seismologists, geodesists with education components for educators and students (orange, bold type). However, since EarthScope data became freely available online in 2004, many other users have discovered and employed EarthScope data (yellow, regular type), ranging from new disciplines using GPS data (e.g., hydrologists) to other facilities (e.g., LIGO) using seismic data. Note that many of the "unanticipated" users are funded by other programs within NSF.*

When EarthScope facility managers were asked about these trends, some have been surprised by the growth in use. Across the EarthScope interviewees, EarthScope administrators, personnel, and users were genuinely excited and stimulated by this growth. However, some sources said that new users, in particular ones who may be unfamiliar with geodetic and seismic data processing or formats (particularly surveyors and utility company representatives), "have had a non-negligible effect on our time" and in some cases, "they require more hand-holding" than the academic geoscience users. In some instances, new users either demand or expect new data products to suit their needs and applications. While these products may have broad applications, facility managers struggle with deciding when to take this on as a facility versus when to "push back" and make this a responsibility of the new users.[72]

By comparison, ARF, NEES, and NOAO/KPNO did not have stories of unanticipated users, and this is likely due in part to the need to physically access these facilities. Even for NEES, which does have a cyberinfrastructure and data freely available online, facility managers did not cite any unanticipated users.

Outside NSF, the HST staff described unanticipated users from the arts and even the fashion industry and remarked that the public availability of the Hubble imagery has led to its use well beyond academic research. The DOE synchrotron facilities have one of the best examples of unanticipated use. When

[72] Similar challenges are faced by Canada's ocean observing networks, Neptune Canada (http://www.neptunecanada.ca/) and Venus (http://venus.uvic.ca/), where new users are demanding new data products, and therefore drawing on more resources than was anticipated from the intended users. Facility managers discussed some of these challenges during a town hall session at the 2012 AGU Ocean Sciences Meeting.

these facilities were created, they were intended largely for the material sciences community. Yet over time, life scientists have became 40% of the users today.[73]

The EarthScope story is an important lesson in the rapid growth of facility use, especially when the facility makes data consistently and freely available online. It is conceivable that facilities built around a cyber-connected networks of distributed sensors such as OOI and NEON may see similar patterns in use over time once they become operational.

5.3 Facility Users and the "Data Tsunami"

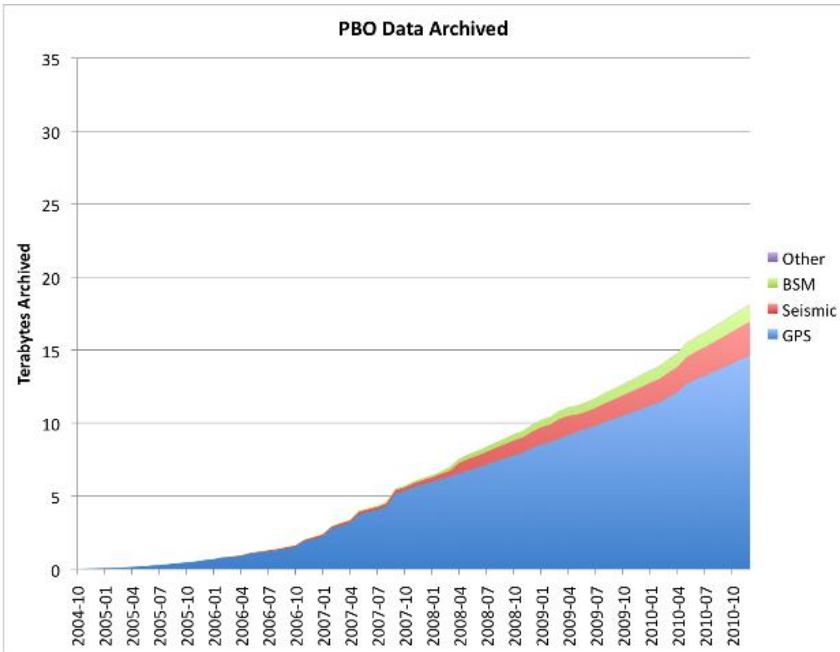
Discussions surrounding the challenges and opportunities presented by "big data" have infiltrated federal science agencies and the private sector.[74] The rapid growth in social media, team science, and interdisciplinary studies has exacerbated the need for coordinated efforts to better understand, store, move, synthesize, and archive large datasets. Federal agencies are responding and recently the US government expressed its commitment to advancing knowledge and building the "big data" workforce with its Big Data Research and Development Initiative.[75]

Big data and lifecycle data management is an important issue for many of NSF's large facilities, which have seen a dramatic increase in data output, and correspondingly, data drawdown. For example, EarthScope began streaming data online in 2004 and has since produced and archived tens of terabytes of data from its different distributed sensors. For the PBO, these volumes have increased at a roughly linear rate. However, as EarthScope PBO data have been used and adopted by new users over time, the terabytes of data *downloaded* have increased exponentially (Figure 13). This "data tsunami" has important implications for sustainable facility use, which is addressed in Chapter 6.

[73] The National Research Council 1999 report on *Cooperative Stewardship* captures the early part of this development and its implications for facility management. See the report's figure 2.1, which shows the increase in the number of life science users from 200 users in 1990 to 1400 users in 1997/8 compared to modest growth in other fields such as material sciences, engineering, and chemistry.

[74] See *The Fourth Paradigm: Data Intensive Scientific Discovery*, available for free electronically at http://research.microsoft.com/en-us/collaboration/fourthparadigm/**.** The 2009 publication is a collection of essays, inspired by Microsoft computer scientist Jim Grey's vision of a "fourth paradigm of scientific discovery" rooted in data-intensive computing and is one of the seminal resources in big data.

[75] The Big Data Research and Development Initiative was announced on March 29, 2012. Led by OSTP in partnership with other agencies including NSF, it is designed to "advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data; harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning; and expand the workforce needed to develop and use Big Data technologies." (http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)



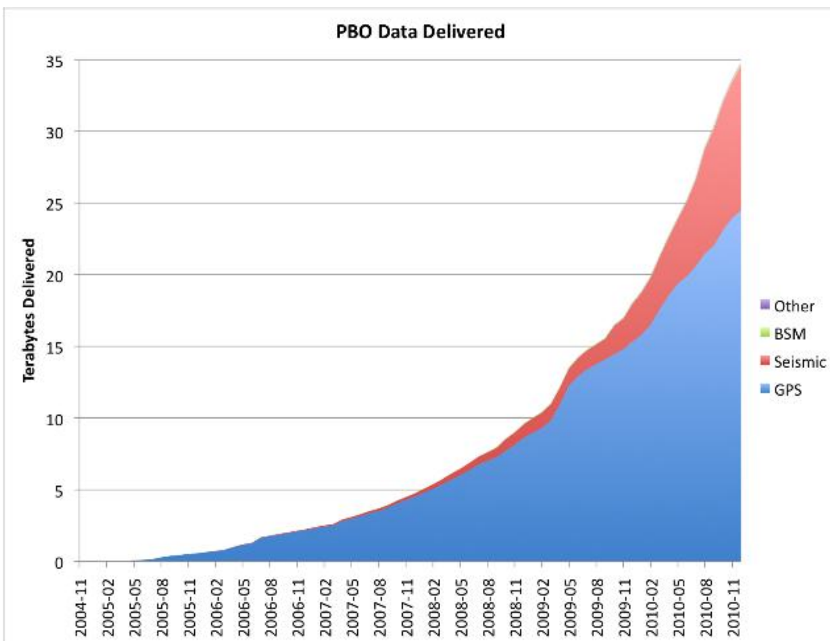


Figure 13: Data Delivery and Drawdown from EarthScope's Plate Boundary Observatory (PBO). *Since it began in 2004, EarthScope has produced and archived PBO data at a near linear rate (top). By comparison, data drawdown (data delivered, bottom) has increased exponentially (courtesy UNAVCO).*

In an increasingly data-intensive research environment, scientists are exploring new ways to exploit archived data. At some facilities such as NASA's Hubble Space Telescope, some users are no longer producing or using new data: instead, they rely entirely on mining archived data for their research. Across the case study facilities, EarthScope and NEES offer the best examples of facilities evolving to accommodate changes in data production, archiving, and use over time.

Example 1: New Views on EarthScope's Cyberinfrastructure

EarthScope has hundreds of users that rely heavily on its streaming data and cyberinfrastructure resources. However, in this study, sources cited shortcomings including the lack of robust visualization tools and data products to make EarthScope data more manageable for a wider audience. One interviewee even stated that in retrospect, it was a mistake to build EarthScope's cyberinfrastructure during the construction phase, because "it was built before the community was really developed." Accordingly, EarthScope's Cyberinfrastructure Subcommittee recently (May 2012) published their "Preliminary Strategic Plan for EarthScope Cyberinfrastructure"[76] to address some of these challenges as EarthScope moves forward. The report describes an enhanced cyberinfrastructure for EarthScope that complements its 2010-2020 Science Plan and is in keeping with the efforts of NSF's new EarthCube initiative.[77] This example shows how a relatively new facility is adapting to changes in technology, user needs, and new funding opportunities surrounding big data since it began operations.

Example 2: NEES Data Grants

Before 2009, a PI could not receive funding to conduct research using only archived NEES data. However, in 2009, NSF made awards available for the sole purpose of NEES data re-use, which has resulted in both an increase in data use, and in community value of NEES' online resources, such as its Project Warehouse. One NEES user said that "people are looking at the data archive as a way to do new research" and felt that "students and younger-faculty see [the Project Warehouse] as a valuable resource."[78] One administrator suspects that NEES data is not yet at a point where it is broadly used, and several sources commented that the earthquake engineering community has and continues to endure "growing pains" in its adjustment to sharing data and working at NEES facilities as opposed to traditional work in their individual labs. However, Hacker et al (2011) have shown the total number of documents, tools, and learning content contributed by users to the NEES Project Warehouse has increased over time (Figure 6), pointing to the community's adoption of data sharing via cyberinfrastructure. While the long-term trends of data sharing and re-use at NEES remain to be seen, many of the NEES managers and users are optimistic about its future use and growth.

Example 3: Beyond NSF - Hubble's Data Miners

During visits at STScI, multiple interviewees described the broad and growing use of the Hubble Space Telescope's (HST) data archive, and pointed to this as one of Hubble's greatest legacies to the field of astronomy. In particular, they described how many users complete their research using archived HST data – where many of these users are never awarded observing time on the HST. The ubiquitous use of the HST archive is illustrated by the increase in the number of science publications using archived data over time (Figure 14). While the total number of publications from observers has remained largely constant, the number of publications using archived data (or a combination of direct observations and archived data) has steadily increased, pointing to the community-wide value of a usable data archive.

---

[76] EarthScope's "Preliminary Strategic Plan for EarthScope Cyberinfrastructure" (May 11, 2012) is online at http://www.earthscope.org/es_doc/highlights/ES_CyberinfratructureStrategicPlan_2012.pdf

[77] http://www.nsf.gov/geo/earthcube/

[78] Other NEES managers had similar observations of the generational differences between engineers who have embraced the NEES cyberinfrastructure and data products versus those who resent it or prefer to not use it.

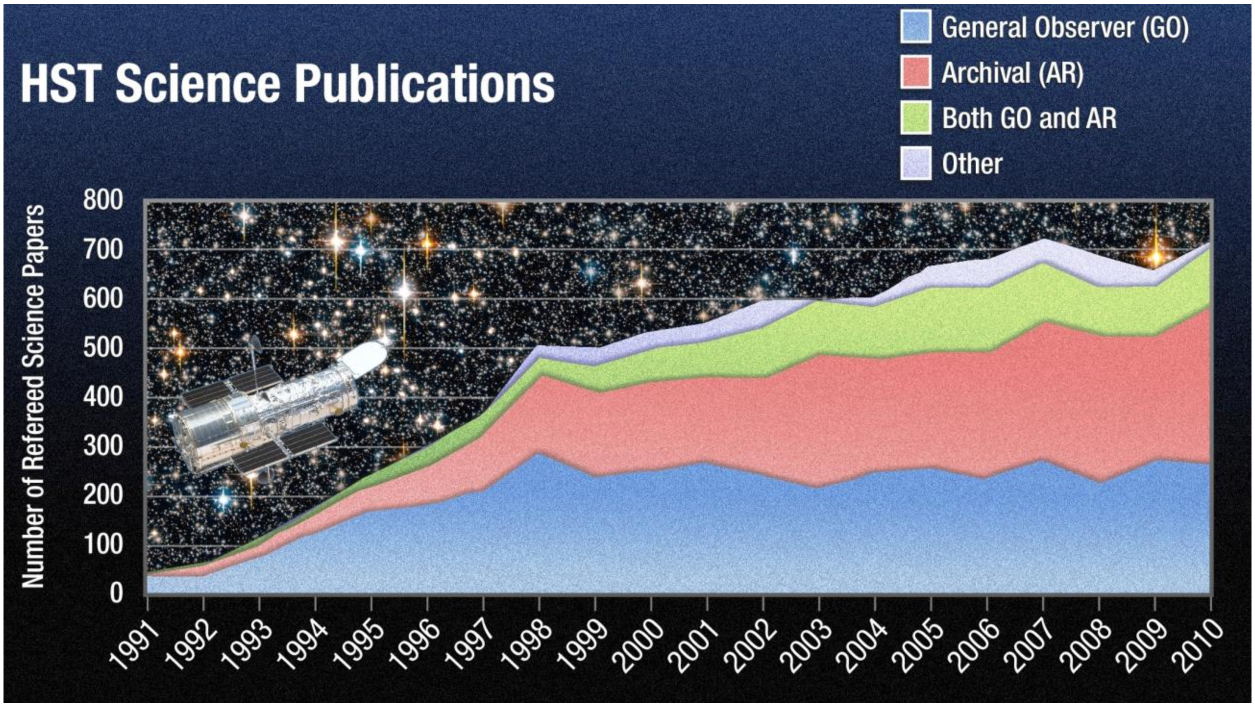


Figure 14: Science publications from the Hubble Space Telescope (HST) since 1991. *Trends show that the total number of publications using archived data has steadily increased over time, while the rate of publications derived from observations has remained relatively constant. (courtesy STScI).*

Clearly, data archives play an important role for large facilities both within and beyond NSF and the increases in data production and data demand will likely increase in the future.[79] This has important implications for lifecycle management of facilities, as discussed in Chapter 6.

5.4 Changes in User Skills and Backgrounds

Interviews at all four of the case studies were asked about how new users and early career scientists were interacting with the facility. Comments about new users varied, but one theme emerged. Of the four case-study facilities, managers and facility staff at EarthScope and NOAO/KPNO explicitly described their observations of changes in user skills and backgrounds in early career scientists who are coming to use their facilities.

From these interviews, early career scientists:

- are more multidisciplinary,
- are more open to collaboration and to sharing data,
- are more data savvy, and
- have higher expectations for the 24/7 availability of data

than their predecessors.

---

[79] The 2009 National Science and Technology Council Report, *Harnessing the Power of Digital Data for Science & Society* describes the need to “ensure that digital scientific data be reliably preserved for maximum use in catalyzing progress in science and society.”

Interestingly, in three separate interviews of administrators at three different facilities (EarthScope, NOAO/KPNO, and NEON), the interviewees (who did not know one another) shared essentially the same story about graduate students at their facilities. One of the interviewees summarized it as, "students these days don't even know how to use a screwdriver- they just expect the data to be there." While this anecdote is amusing, it points to changes in how research is being conducted in a much more data-intensive and socially-connected environment than was available even five years ago. These observations have important implications for the management and planning of facilities in the area of workforce development, which is addressed in Section 6.4.

Outside NSF, staff at STScI commented that many of the early career users of the Multimission Archive (http://archive.stsci.edu/) are more adept at querying large datasets using programming. They have observed users evolve from searching for specific, known datasets by hand to writing programs using SQL or other programming languages to automatically search and mine the available databases.

In contrast, sources at DOE's Office of Science did not report any changes in their user background, skills, or expectations for the BES facilities. It is important to recognize that for these facilities, experimental data are not available to the community and DOE does not maintain a data archive for experiments conducted on the BES instruments.

5.5 Discussion

This chapter investigates how facility use changes over time. Results show that new users discover NSF's large facilities through a variety of formal and informal pathways. The degree to which a facility "actively seeks" new users varies with facility type, activity, and goals and different facilities identify and seek new users through different means. Over longer time scales, facilities may experience changes in use over time, where new users from new fields of research or other sectors adopt the use of the facility. This change can open new areas of research, but can also impact (and sometimes strain) facility resources, as shown by the example from EarthScope.

Most recently, many of NSF's facilities are facing a "data tsunami" and are grappling with the challenge of storing and archiving unprecedented amounts of data. All of the case study facilities are working to develop new data products, data analysis tools and new ways to embrace the exciting new landscape of big data. As a comparison, the HST has already shown the advantages of data re-use and has been able to maximize scientific return on careful data archiving[80] and management, as shown by the consistent increase in publications using archived data.

The era of big data also appears to have consequences for some of the new users interacting with facilities such as NOAO/KPNO and EarthScope, where new/early career scientists have higher expectations for the availability and utility of data than their predecessors. As shown in the next chapter, this observation has implications for the role of large facilities in scientific workforce development.

[80] Examples from biotechnology also point to the positive scientific return of data archiving. See the article "Data archiving is a good investment," in the 19 May 2011 issue of *Nature* and links therein for additional examples.

6. Applying User Characterization to Facility Management and Planning

The previous four chapters have established a conceptual framework for analyzing large facility users and facility use; explored facility users by the numbers through user tracking; examined facility use through the lens of user feedback to facilities; and investigated the changes in facility users and use over time. How can this information be applied to managing existing facilities and planning future facilities? This chapter focuses on four topics of facility management that are directly related to facility use: program evaluation, facility mission, identifying areas of synergy, and workforce development.

6.1 Facility Utilization, Metrics, and Program Evaluation

Characterizing and quantifying large facility use will likely become increasingly important as the federal government continues to focus on developing metrics and evaluation tools for assessing its investments in science and engineering research in an era of reduced federal funding.

Across the US government, numerous efforts are underway to create new databases and analysis tools to enable funding agencies and policy makers better synthesize, understand, and defend their investments in scientific and engineering research. This is largely driven by the decreasing availability of federal funds for science research[81] and the government's desire to better understand the drivers of innovation[82] and economic competitiveness. New developments in evaluation and research tracking efforts include programs like the Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science (STAR Metrics), which is "a federal and research institution collaboration to create a repository of data and tools that will be useful to assess the impact of federal R&D investments."[83] Other efforts include the recently-announced Open Researcher and Contributor ID (ORCID) system, which will work like a barcode or social security number, tagging each of the world's scientists with a unique identification number that will connect scientist's names to their publications, grants, citations, and contact information.[84] Yet another example is the Sci^2 Tool,[85] which has been developed to analyze and visualize scientists' professional networks using federal award, publication, and patent databases. As a final example (of many more), NSF's Science of Science and Innovation Policy (SciSIP) initiative, which was created in 2008 to fund "researchers from all of the social, behavioral and economic sciences as well as those working in domain-specific applications such as chemistry, biology, physics, or nanotechnology" to "advance the scientific basis of science and innovation policy. Research funded by the program thus develops, improves and expands models, analytical tools, data and metrics that can be applied in the science policy decision making process."[86]

[81] A May 18, 2012 memo (M-12-14) from the Office of Management and Budget (OMB) states, "since taking office, the President has emphasized the need to use evidence and rigorous evaluation in budget, management, and policy decisions to make government work effectively. This need has only grown in the current fiscal environment…. Agencies should demonstrate the use of evidence throughout their Fiscal Year (FY) 2014 budget submissions."

[82] See NSTC 2008 report and the book, *The Science of Science Policy: A Handbook*

[83] STAR Metrics was created in 2010 and has been led by NSF and the National Institutes of Health (NIH) under the guidance of OSTP. Additional information about STAR Metrics and its origins and goals are available at http://www.starmetrics.nih.gov/.

[84] See the article by Butler in the May 31, 2012 issue of *Nature*.

[85] The Sci^2 Tool is available for free at http://sci2.cns.iu.edu/. Online tutorials guide new users in its applications and the developers of the tool often give day-long training sessions at conferences and universities.

[86] Information on NSF's SciSIP program, including current funding opportunities, is available at http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501084.

However, research using these tools and databases to quantify science research and its outputs at NSF has so far have been limited to testing specific programs and/or division portfolios. Exactly how and where large facilities and their users fit into these efforts is unclear and is currently under discussion at the agency. But, given the growing emphasis on using evidence and program evaluation for setting national research priorities and budgets,[87] it is likely that this will become increasingly important for facilities in the future, with facility utilization data being an essential component of evaluation and assessment.

6.2 Maintaining Facility Mission through Changes in Facility Use

Facility users and the type of use may change as science and technology change over time. It is important for facilities to recognize opportunities for growth and balance these with their mission.

Of all the case study facilities, EarthScope provides the best example of how a facility built by and for a select community of users (i.e., geologists, seismologists, geodesists) has expanded to include many additional users from other disciplines (Figure 12). While this expansion is exciting and is an outstanding demonstration of the broad scientific and commercial utility of GPS and seismic sensing data, how does a facility like EarthScope prioritize its resources to ensure first-rate support for both the intended users and the unanticipated users? If the new, "unintended" users demand training or new data products, does the responsibility for creating these fall to the new users, or is it absorbed by the facility? EarthScope facility staff claimed that some of the new users "have a non-negligible effect on our time" and that new employees were hired to accommodate the growing demands for data processing. Clearly, unanticipated users hold great promise for advancing science, but can strain a facility's resources. Other case study facilities did not report "unanticipated" users – suggesting this may be a trend unique to facilities that are either enabled by a CI, and/or that have open and freely-available data. Facilities and funding agencies will likely find the combination of user tracking and informal feedback to be important tools for understanding and responding to shifts in community needs throughout the lifecycle of a facility.

6.3 Identifying Areas of Synergy among Facilities

Trends and observations in facility utilization across NSF can point to areas of synergy and possible new avenues of collaboration between facilities, centers, and initiatives that may otherwise go unseen.

Recently, there have been requests from Congress and the National Science Board (NSB) to identify areas of synergy across facilities. The America COMPETES Reauthorization Act of 2010 specifically calls on NSF to "coordinate and collaborate with other Federal agencies, including the DOE's Office of Science, to ensure that joint investments be made where practicable;" and, "for facilities in which multiple disciplines will be possible, the Director should include multiple units within the Foundation during the planning process." During the February, 2012 National Science Board (NSB) meeting, the Board requested information on what interconnections exist between NSF's operational facilities (Coles, 2012).

Examining how facilities are being used – and how researchers intend to use facilities in the future – can be an effective tool for identifying synergies amongst facilities. As scientific research becomes increasingly interdisciplinary,[88] some facilities are seeing an increase in users taking advantage of different facilities and resources to accomplish their work. This is particularly true in fields like astronomy, where astronomers rely on a "system" of privately- and federally-funded telescopes to answer increasingly complex questions. Jannuzi and Valenti (2012) show this explicitly in the results of their 2011 survey of the ground-based optical and infrared astronomy community (Figure 10). In their report to the

[87] See OMB Memoranda M-12-14 and M-12-15
[88] http://chronicle.com/article/National-Science-Foundation/130757/

NSF Astronomy Portfolio Review Committee,[89] they state, “the vast majority of researchers (more than 80%) heavily use numerous and diverse facilities because no single capability can provide the range of data required.” To further illustrate the inter-reliability of astronomy’s “system” of facilities, a 2011 astronomy article[90] published in *Science* had 63 authors who used data from multiple telescopes including the W.M. Keck Observatory, the Gemini Observatory, the Chandra X-ray Observatory, the UK Infrared Telescope, and many others. Data from these different facilities provided an unprecedented multi-wavelength view of a gamma ray burst, leading to a high impact publication.

Although astronomy is a more mature field than many others, “systems” of facilities in other disciplines are becoming increasingly important. In the geosciences, researchers who are planning work using the Ocean Observatories Initiative’s (OOI, currently in construction) network of sensors rely on ARF vessels for deployment and maintenance of instruments. Data managers at the R2R Data Repository remarked that one of the top geographic areas for data requests is the northeast Pacific, which is the future site of the OOI’s Regional Scaled Node.[91] Researchers in the Integrated Ocean Drilling Program (IODP) are developing and testing borehole sensors that may ultimately be connected to the OOI to transmit data in real time. Geologists have recently begun exploring the field of seafloor geodesy,[92] and how the precedent for terrestrial continuous GPS data (set by EarthScope’s PBO) could be translated to the seafloor. In addition to these overlaps, NSF’s EarthCube initiative,[93] funded jointly by the Geosciences directorate and the Office of Cyberinfrastructure, envisions combining data from all of these efforts (and more) into a searchable, usable database for geoscience research. These will inevitably lead to increased interconnections between facilities, allowing users to interchangeably use information from different facilities for their work.

From a management perspective, there may be ways to expand utilization studies to elucidate how facilities can share and/or leverage resources. For example, can facilities that support research in similar disciplines support one another’s education and public outreach efforts? Or, could facilities that pioneer new ways of managing, archiving, and distributing data use one another’s mechanisms as building blocks? Although each facility and the respective communities have their own needs, it is conceivable that by studying use and user-facility interactions with specific goals in mind such as finding areas of overlap or synergy in cyberinfrastructure or education, could be useful for both facility planning and management.

6.4 Workforce Development

Trends in changes in new user skills, backgrounds, and expectations are important indicators of future needs for workforce development and user training.

Research programs and some facilities are seeing early career scientists who are more data savvy and prone to collaboration. These observations are evident across science disciplines and extend beyond NSF’s large facilities.[94] The rise of team science[95] also presents both a challenge and opportunity to

---

[89] Available online at http://ast.noao.edu/sites/default/files/SystemRoadMapCommittee_PRCsubmission.pdf
[90] See Levan et al., 2011, *Science* v. 333, p. 199-202.
[91] See http://www.interactiveoceans.washington.edu/story/The+OOI+RSN+Cable+System for more information and a detailed description.
[92] This was a session topic at the 2012 UNAVCO Science Workshop – see http://www.unavco.org/community/meetings-events/2012/sciworkshop12/draft_agenda.pdf
[93] See http://earthcube.ning.com/
[94] A perfect example is illustrated in a recent *Science* “NextGen Voices” article where early career scientists weighed in on their view of how will the practice of science change in their lifetime. Comments

facilities and science workforce development. One of the hallmarks of many big or long-term science projects is research conducted by large groups, or teams of scientists and engineers working across disciplinary and often international boundaries to solve increasingly complex grand challenges. Combining team science with social media, more collaborative research environments, and open data, what will facility research look like in the future? Will doctoral dissertations be completed by teams of graduate students? Will new collaborations focused on data mining uncover new fields of research or new applications for old data? How will early career scientists be recognized for their contributions to team science projects or large datasets?[96] How might academic institutions reorganize themselves to reflect the changes in ways that science is being done and how would this affect degree requirements?

It is important to provide new users and early career scientists with the tools they need to successfully work in a more data-intensive environment. Industry recognizes this[97] and in science, it will become more imperative as research becomes increasingly computational and integrated across disciplines. In response to these trends, some facility managers feel that facilities should take a central role in becoming a hub for training "data science." While efforts in this arena are underway in some fields,[98] representatives from two of the case study facilities expressed concern about the lack of training opportunities for current science users to learn how to effectively use large datasets for their research. One administrator went as far as saying "we are doing our community a disservice if we don't provide this training to our students." Whether or not it is best to assign this responsibility to facilities or to other entities that can reach more students across the discipline is unclear. However, because many facilities have high visibility within their fields and typically have complex data management programs in place, it is conceivable that they could play a central role in creating and/or supporting data training programs.

Finally, it is also important to recognize large facilities as rich "training grounds" for many users who are early career scientists. Facilities offer a unique combination of resources that cannot be found at individual university labs: undergraduates, graduate students and post-doctoral scholars at facilities have access to expert technical and support staff, world-class equipment and instrumentation, sophisticated data analysis tools, and the opportunity to interact with peers from other institutions. Facilities like NOAO/KPNO are unique in providing instrument time and equipment to students. For optical astronomers, earning time on the KPNO telescopes is "easier" in comparison to time on privately-funded telescopes, and there are more opportunities for students to work on the KPNO telescopes than on others

---

ranged from "we are… entering the Era of Interdisciplinarity" to "a new career path in science will emerge: the professional scientific data manager will have a unique skill set from areas such as statistics, large database administration, and information design" and "the key theme distinguishing the future practice of science will be integration…. collaboration will be critical for scientific discovery." (Sills, 2012).

[95] There are many resources documenting this trend – as a start, see Wuchty et al, "The Increasing Dominance of Teams in Production of Knowledge" in *Science* 316, 1036 (2007).

[96] The book *Reinventing Discovery: The New Era of Networked Science* by Michael Nielsen addresses this and urges academia to explore new ways for incentivizing contributions to and the use of large datasets and data analysis tools (see p. 182). In general, this text provides an thought-provoking overview of the current state of big data, team science, and the "new era of networked science."

[97] See http://insightdatascience.com/fellowship.html for an excellent example of how the private sector is training new data scientists to work in Silicon Valley.

[98] The annual Cyberinfrastructure for Geoscientists Summer Institute began in 2004 and is funded by NSF. The week-long course is taught by geoinformatics researchers and practitioners and includes both introductory material as well as technical training in data mining and cloud computing. Some of the instructors are involved in NSF's EarthCube initiative. Additional information is at http://www.geongrid.org/index.php/education/summer_institute/. In a similar vein, the University of New Mexico (home of DataONE, http://www.dataone.org/) has developed a graduate courses in data analysis and visualization as part of the Walter E. Dean Environmental Information Management Institute (http://elibrary.unm.edu/courses/eimi/index.php).

in the O/IR observatory system. For the ARF vessels, graduate students gain invaluable experience participating in research expeditions, and this experience often forms the foundation for their future work as PIs later in their careers. The role of facilities in training the next generation of scientists should not be overlooked when considering future plans for scientific workforce development.

7. Recommendations and a Roadmap for Future Work

Facility utilization analysis is ripe for many areas of application and future growth. With additional resources, comprehensive user analysis could lead to a deeper understanding of users and facility use across NSF facilities and results could inform mid-to long-term efforts in developing metrics for evaluating facility utilization[99] and facility lifecycle planning. The following discussion presents recommendations for implementing the findings of this study into facility management and ideas for future work in analyzing facility utilization using this study as a foundation.

7.1 Observations and Recommendations

Summarized below are three key observations and recommendations from this study and suggestions for near- to mid-term implementation.

Observation 1: Data on users is variable and is not collected in a standardized way across facilities.

Recommendation: Explore the feasibility of creating common standards and/or identifying best practices for tracking and reporting facility use and investigate how these data could be integrated with data collected across NSF.

Implementation:

- Consider creating an internal working group of representatives from multiple NSF facilities to compare how users are tracked; how this information is used; and what pieces of additional information would be most useful to facility management. A summary of these findings could be presented and discussed at either a Large Facility Program Officer's Forum[100] and/or at the annual Large Facilities Workshop to collect additional input from facility management representatives.
- To capitalize on data that already exists on facility use, consider working towards identifying a minimum set of usage metrics for every facility to report in their annual report or annual work plan (one common place for all facilities). This should be flexible to allow for the differences between facilities. It is conceivable that this could be a set of "if-then" statements: for example, if the facility provides open access to data online, then the facility can report # of website visits or quantities of data downloaded.
- Initiate conversations with colleagues in the Office of Integrative Affairs (OIA) who are working on database needs across the Foundation to consider including facility usage in NSF database development.

Observation 2: Some facilities see changes in use over time, which may include unanticipated users.

Recommendation: Explore new tools for systematically examining, documenting, and evaluating changes in facility use over time.

Implementation:

- Consider encouraging periodic (e.g., every 3-5 years) community surveys as a tool to collect information across facility- and disciplinary- boundaries to assess community demand for and use of "systems" of facilities.

---

[99] This is specifically identified as a mid-term action in NSF's Strategic Plan for FY2011-2016.

[100] These are internal seminars held monthly at NSF and are organized by the Large Facilities Office.

- Use topic modeling of both awarded and declined proposals to examine changes over time in community interest and use of the facilities. Compare results to information from community surveys to assess changes in facility use.
- Encourage facility review committees to identify, examine, and report on any changes in use during facility operations reviews.

Observation 3: CI-enabled facilities are seeing more sophisticated users exploiting open data.

Recommendation: Examine the role of facilities in the era of big data and evaluate the impact of this change on workforce needs.

Implementation:

- Determine facility workforce needs for supporting users working in a more data-intensive research environment by first initiating conversations among facility managers and users. This could be a breakout group topic at the 2013 Large Facilities Workshop. It could also be posted on the NSF Large Facilities Program Management Community of Practice LinkedIn group site[101] to gather additional input.
- For examining scientific workforce needs, it may be useful to work with the Office of Cyberinfrastructure (OCI) and program managers across NSF to document existing training opportunities/workshops in data analysis both within and outside facilities. This information may be used to identify areas where facilities could either help with and/or benefit from ongoing efforts in this area.
- Encourage facilities to examine how education and public outreach efforts can effectively communicate the use of open data in scientific research and emerging career opportunities in data intensive science.

7.2. Recommendations for Future Work in Facility Utilization Analysis

Expanding User Analyses to Include More Facilities

In this study, four facilities were selected as case studies by necessity to hone the scope of this project. However, as discussed in Chapters 1 and 2, the diversity of NSF's large facility portfolio is high: each facility is unique and has its own ways of interacting with its user community. Characterizing facility utilization at an NSF-wide scale would greatly benefit from broadening the scope of this study to include at a minimum one to two physics facilities, and/or ultimately expanding to include all of NSF's operational large facilities to find true commonalties across NSF's portfolio.

Examining International Users

In addition, the international community of users is not addressed in this study. Although international users would likely fall into the seven "types" of users identified in Section 2.2, it would be helpful to compare NSF's investments in domestic and international facility users to elucidate who, on an a collective international level, is benefitting from NSF's investments in science infrastructure. Facilities that rely heavily on international collaborations such as IODP, the Large Hadron Collider (LHC), and the Gemini Observatory would be ideal candidates for including in a comparative analysis to examine international users. Results may highlight new avenues for building international partnerships,[102] and/or may provide examples of successful collaborations and lessons learned that can be applied to future facilities.

---

[101] To join, one must have a registered login on LinkedIn and be approved by the group manager.

[102] Increasing international partnerships and collaborations is part of NSF's Strategic Plan for FY2011-2016.

Building on Utilization Analyses to Investigate Education and Public Outreach
The educator and public users are not well-studied in this work. It is clear that facilities play an important role in STEM education and public engagement – but how do these efforts compare to those funded by other programs at NSF? How should facilities prioritize their education and public outreach endeavors? What are the best ways to engage the user community in education and public outreach efforts? Are there ways that facilities supported by the same directorate (e.g., ARF, EarthScope, IODP, IRIS, NCAR and OOI in the Geosciences Directorate), and/or facilities that support similar disciplines (e.g., EarthScope, IRIS, and NEES) can leverage one another's efforts in this arena? How should education and public users be incorporated into assessing the vitality of a facility? Although NSF's large facilities are designed to fulfill the needs of the scientific research community, the NSF is committed to science education and: analyses of how educators and the public interact with a facility would be an important asset in working towards answering some of these questions.

Employing Community Surveys as an Analysis Tool
Future studies in facility user analysis may be able to take advantage of online survey technology to collect information for informing answers to questions such as "what is the definition of a facility user" and/or to more systematically identify quantitative user data collected across facilities. One of the challenges in this study was collecting this information primarily through interviews and site visits from a sampling of facility representatives. For future work, especially across more than four facilities, an online survey (designed using a common survey tool such as SurveyMonkey.com, and sent directly to a known list of unique respondents representing facility management and users – or professionally executed by an organization such as STPI or RAND) would be a valuable tool for systematically collecting a rich amount of information in a short period of time. It is important to recognize that although survey responses can be gathered relatively quickly, surveys cannot replace interviews. The author found that the interview process provided invaluable anecdotal information, historical insights, and leads to new resources that typically arose organically, and that may not have been revealed during a survey or written questionnaire.

It should be noted that at the NUFO Annual Meeting in June, 2012,[103] the question of "what is the definition of a user" came up in multiple sessions. As a result of these discussions, NUFO's Administrative Affairs Working Group[104] resolved to create and distribute a survey of the NUFO community to formulate a community-informed definition of users for NUFO. Results from the NUFO survey will undoubtedly be a useful comparison to results from this work and future studies on facility utilization.

Network Mapping and Integrating User Tracking Data
Integrating user tracking data with network mapping could provide powerful assessments of facility use. Network (or community) mapping is a tool that is increasingly being used in the science of science policy and the science of team science arenas to integrate and visualize information across existing networks of researchers and institutions. Mapping can be used to facilitate finding collaborators or reviewers, connecting students with projects, forging connections between academia and industry, engaging the public, and providing policy makers with decision analysis tools. Recent work in network analyses has focused on interpreting individual PI and/or authorship networks, knowledge flow patterns, funding patterns of different agencies, and the relationships between scientist networks and patents or high profile prizes.[105] Many science network research programs such as STAR Metrics and VIVO[106] are using these

[103] See http://lansce.lanl.gov/NUFO/agenda.shtml
[104] See http://www.nufo.org/organization/administrative_affairs.aspx for additional information.
[105] The Science of Science (Sci^2) Tool (https://sci2.cns.iu.edu/) is one of several open source software packages that is available for creating network maps. Their website has many useful online tutorials and presentations: the opening presentation on the "Science of Science Research and Tools"

techniques to evaluate funding portfolios, to identify reviewers, and to better understand the national innovation ecosystem.[107] Most of these programs are employing existing databases (such as funding databases, patent registries, and bibliometric databases such as Web of Science and Scopus) to inform these maps.[108] However, because many of these databases are decentralized, there is significant energy being put towards integrating these datasets into central databases that can be queried.[109] In the future, it may be possible to integrate facility user data from different user tracking methods (such as data downloads, users on site, and registered online users) with these datasets to produce comprehensive network maps for analyzing and assessing facility utilization.

With integrated databases at NSF, network maps could serve different facility analysis purposes. For example, community maps could show how the geographic distribution of fixed-site facility (e.g., observatories or labs) users compares to CI-enabled facility users. Or, similar to Figure 10, community maps could be used to examine a discipline's use of different research facilities (across NSF, and/or including other facilities). Finally, network maps that integrate award data with user names, user types, user home institutions, and facilities could show the relative distribution of users who receive funding from multiple sources to use national facilities. This would be particularly informative for examining facility use across the Foundation and for comparing domestic and international use of facilities.

Applications Beyond Large Facilities

Both the methods and the conceptual framework for studying facility utilization described in this report are scalable and could be applied to other multi-user resources supported by NSF, such as Science and Technology Centers[110] and/or mid-scale infrastructure.[111]

(http://ivl.cns.iu.edu/km/pres/10-NIH-Tutorial-01.pdf) provides a useful overview (with many illustrations) of the variety of applications of this powerful software.

[106] From VIVO's website, "VIVO is an open source semantic web application originally developed and implemented at Cornell University. When installed and populated with researcher interests, activities, and accomplishments, VIVO enables the discovery of research and scholarship across disciplines at that institution." See http://vivoweb.org/ for additional information.

[107] This is addressed throughout *The Science of Science Policy: A Handbook*, published in 2011. In particular, Section Two of this text provides valuable insights into "Empirical Science Policy – Measurement and Data Issues."

[108] The Sci^2 Tool, for example, lists available databases here: http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/8.1+Datasets

[109] As an example, see chapter 12 of *The Science of Science Policy: A Handbook,* which includes a table of science and technology databases and their locations.

[110] See http://www.nsf.gov/od/oia/programs/stc/ for a program description and list of centers.

[111] See the National Science Board's May 4, 2012 report on "The NSF Support of Unsolicited Mid-Scale Research" (http://nsf.gov/pubs/2012/nsb1222/nsb1222.pdf) for a discussion of mid-scale research.

8. Conclusions

This study presents a case study analysis of users and facility utilization to determine the definition of a user, who uses NSF's facilities, how users access facility resources, and how use changes over time. Observations are used to define how lessons learned from facility utilization analysis can be applied to facility management and planning. This work establishes a critical framework for addressing facility utilization and shows multiple avenues for exploring facility utilization in the future to further examine facility use across NSF's diverse large facility portfolio.

At the outset of this study, several interviewees claimed they knew who their facility users were while others remarked that they didn't really know who is using their facility on a collective level, highlighting the difference between NSF's large facilities and the perception of facility utilization. This study shows that NSF's large facilities are serving a very broad range of users across the research, education, and in some cases, private sectors. Analyzing facility utilization – through interviews, quantitative user tracking data, surveys, and user feedback to facility management – is key to understanding who is benefitting from NSF's investments in large facilities, what synergies exist among operational facilities, and identifying new areas of research and future needs for science workforce development for NSF large facilities and will likely become increasingly important in the future.

REFERENCES

Apai, D., J. Lagerstrom, I.N. Reid, K.L Levay, E. Fraser, A. Nota, and E. Hennekin (2010) Lessons from a High-Impact Observatory: The Hubble Space Telescope's Science Productivity between 1998 and 2008, *Publications of the Astronomical Society of the Pacific*, vol. 122, p. 808.

Berman, F. (February 18, 2005). "SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences," San Diego Supercomputer Center. Retrieved May 31, 2012.

Butler, D. 2012 "Scientists: Your Number is Up," *Nature*, vol.485, p. 564. (31 May, 2012)

Coles, M. *2012 NSF Facility Plan*, presented to the National Science Board, February 2, 2012 at NSF, Arlington, VA.

Corona Insights (2010) *Stakeholder Study: National Ecological Observatory Network.* Report prepared by Corona Insights for NEON, Inc.

DOE Office of Science Scientific User Facilities Division website, http://science.energ.gov/bes/suf

Department of Energy, *User Facilities of the Office of Basic Energy Sciences: A National Resource for Scientific Research* (US DOE Office of Science, 2009) http://science.energy.gov/~/media/bes/suf/pdf/BES_Facilities.pdf and http://science.energy.gov/~/media/bes/suf/pdf/BES_Facilities_Number_of_Users.pdf

Fealing, K.H., Lane, J., Marburger III, J., and S. S. Shipp, eds. (2011) *The Science of Science Policy: A Handbook*, Stanford University Press, Stanford, CA.

Hey, T., S. Tansley, and K. Tolle, eds**.** (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Corporation. (Available online at http://research.microsoft.com/en-us/collaboration/fourthparadigm/)

Hacker, T.J., I. Mathew, G. Rodgers, and S. Pejsa, (2012) Data Cyberinfrastructure: Management of Large-Scale Scientific Data, *submitted to Elsevier*.

Hacker, T.J. and A.J. Magana (2011) A Framework for Measuring the Impact and Effectiveness of the NEES Cyberinfrastructure for Earthquake Engineering, *available online at* http://nees.org/resources/3964/download/NEES_CI_Metrics_NEEShub.pdf.

Januzzi, B. and Valenti (2012) Ground-based O/IR System Roadmap Committee Community Survey: Summary of Results from US Based/Sponsored Respondents., v1.5, http://ast.noao.edu/sites/default/files/SummaryDocumentSystemRoadmapCommunitySurveyV1.5.pdf (retrieved May 31, 2012)

Levan, A.J. et al. (2011) An Extremely Luminous Panchromatic Outburst from the Nucleus of a Distant Galaxy, *Science*, 333, 199-202.

Meth, C. and K. Ludwig (2012) "IODP: Engaging New Stakeholders While Maintaining Established Relationships," *presented at AGU Ocean Sciences meeting*, February 20, 2012, Abstract 12864.

Miller, William L. *Preconstruction Planning for Large Science Infrastructure Projects: A Comparative Analysis of Practices and Challenges at DOE, NASA, and NSF*, Report to NSF, January, 2010.

National Research Council. *Cooperative Stewardship: Managing the Nation's Multidisciplinary User Facilities for Research with Synchrotron Radiation, Neutrons, and High Magnetic Fields*. Washington, DC: The National Academies Press, 1999.

National Research Council. *Setting Priorities for Large Research Facility Projects supported by the National Science Foundation*. Washington, DC: The National Academies Press, 2005.

National Research Council. *Portals to the Universe: The NASA Astronomy Science Centers.* Washington, DC: The National Academies Press, 2007.

National Research Council. *Science at Sea: Meeting Future Oceanographic Goals with a Robust Academic Research Fleet.* Washington, DC: The National Academies Press, 2009.

National Science Foundation, *FY2011 Budget Request to Congress*, 2010, http://www.nsf.gov/about/budget/fy2011/

National Science Foundation, *Large Facilities Manual* (NSF 10-12), 2011, http://www.nsf.gov/pubs/2010/nsf10012/nsf10012.pdf

National Science Foundation, *Empowering the Nation Through Discovery and Innovation: National Science Foundation Strategic Plan for Fiscal Years (FY) 2011-2016* (NSF 11-047), 2011, http://www.nsf.gov/news/strategicplan/nsfstrategicplan_2011_2016.pdf

National Science Foundation, Grant Proposal Guide, document gpg11001/NSF 11-1, updated January 2011. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_index.jsp.

National Science Foundation, “Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans,” press release 10-077, published May 10, 2010.

National Science and Technology Council, *Harnessing the Power of Digital Data for Science and Society,* Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council, January, 2009.

National Science and Technology Council, *The Science of Science Policy: A Federal Research Roadmap* Report on the Science of Science Policy to the Subcommittee on Social, Behavioral and Economic Sciences Committee on Science, November, 2008.

Nielsen, Michael, *Reinventing Discovery: The New Era of Networked Science*, Princeton: Princeton University Press, 2012.

Office of Management and Budget (2012): “Use of Evidence and Evaluation in the 2014 Budget” Memorandum M-12-14. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-14.pdf

Office of Management and Budget (2012): “Science and Technology Priorities for the FY 2014 Budget” Memorandum M-12-15. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-15.pdf

Piwowar, H.A. Vision, T.J., and M.C. Whitlock. (2011) Data archiving is a good investment, *Nature*, 473, p. 285.

Sci^2 Team. (2009). Science of Science (Sci^2) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu.

Sills, J., ed. (2012) “NextGen VOICES Results: Future of a Generation” *Science*, 335, 36-38. (6 January, 2012)

White House Office of Science and Technology Policy, 2012, Press Release: Obama Administration Unveils “Big Data” Initiative: Announces $200 million in new R&D Investments, March 29. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.

Wuchty, S., B.F. Jones, and B. Uzzi (2007) “The Increasing Dominance of Teams in Production of Knowledge,” *Science* 316, p.1036-1039.

Zhang, J., M.S. Vogeley, and C. Chen (2010) “Scientometrics of big science: a case study of research in the Sloan Digital Sky Survey, *Scientometrics*, 86, p. 1-14.

Zimmerman, A. (2011) NEES User Survey: Final Report, *submitted June 28, 2011 to NEEScomm*.

Zhao, J.,C. Chen, and M.S. Vogeley. (2011) Modeling Users’ Data Usage Experiences from Scientific Literature. From A. Marcus (Ed).: Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 337-346. Springer-Verlag, Berlin.

APPENDICES

APPENDIX A: NSF'S LARGE FACILITIES PORTFOLIO

NSF has 19 facilities in operation; six in construction, and more in planning, as illustrated below. NSF does not manage or operate any of its facilities directly – this responsibility is contracted to awardee institutions which include universities, companies, and consortia of universities and/or academic institutions.

Table A-1: NSF's Large Facilities

Directorate	Division	Facility	Acronym	Status	Awardee
Biological Sciences (BIO)	Office of the Asst. Director (OAD)	National Ecological Observatory Network	NEON	construction	NEON, Inc.
Engineering (ENG)	Civil, Mechanical and Manufacturing Innovation (CMMI)	George E. Brown, Jr. Network for Earthquake Engineering Simulation	NEES	operation	Purdue University
	Electrical, Communciations & Cyber Systems (ECCS)	National Nanotechnology Infrastructure Network	NNIN	operation	13 university consortium, with network offices at Cornell U. and Stanford U.
Geosciences (GEO)	Atmospheric and Geospace Sciences (AGS)	National Center for Atmospheric Research	NCAR	operation	UCAR
	Earth Sciences (EAR)	EarthScope		operation	UNAVCO/IRIS
		Incorporated Research Institutes for Seismology	IRIS	operation	IRIS
	Ocean Sciences (OCE)	Academic Research Fleet	ARF	operation	UNOLS
		Integrated Ocean Drilling Program	IODP	operation	IODP-MI
		Alaska Region Research Vessel (ARRV)	ARRV	construction	U. of Alaska - Fairbanks
		Ocean Observatories Initiative	OOI	construction	Cons. for Ocean Leadership
Math & Physical Sciences (MPS)	Astronomical Sciences (AST)	Arecibo Observatory (formerly National Astronomy and Ionosphere Center)	(NAIC)	operation	SRI International
		Gemini Observatory	Gemini	operation	AURA
		National Optical Astronomy Observatory	NOAO	operation	AURA
		National Radio Astronomy Observatory	NRAO	operation	AUI
		National Solar Observatory	NSO	operation	AURA
		Advanced Technology Solar Telescope	ATST	construction	AURA
		Atacama Large Millimeter Array	ALMA	construction	AUI
		Large Synoptic Survey Telescope	LSST	planning	AURA
	Materials Research (DMR)	Cornell High Energy Synchrotron Source	CHESS	operation	Cornell U.
		National High Magnetic Field Lab	NHMFL	operation	Florida State U., U. of Florida, Los Alamos Ntnl Lab
	Physics (PHY)	IceCube Neutrino Observatory	IceCube	operation	U. of Wisconsin
		Large Interferometer Gravity-wave Observatory	LIGO	operation	Caltech
		Large Hadron Collider ATLAS and CMS Detectors	LHC	operation	
		National Superconducting Cyclotron Laboratory	NSCL	operation	Michigan State U.
		Advanced LIGO	Adv-LIGO	construction	Caltech
Office of the Director (OD)	Office of Polar Programs (OPP)	U.S. Antarctic Program	USAP	operation	NSF

APPENDIX B: DETAILS ON METHODS AND DATA COLLECTION

Facility Categorization

To select cases, the 19 operational facilities were grouped by multiple categories to identify similarities and differences across the portfolio (Table B-1).

Case Study Facilities

Four case study facilities were selected for this study: the Academic Research Fleet (ARF), EarthScope, the Network for Earthquake Engineering Simulation (NEES), and the National Optical Astronomy Observatory (NOAO). Brief descriptions[112] of each facility are below.

1. **Academic Research Fleet:** The Academic Research Fleet consists of 21 ocean research vessels in the University-National Oceanographic Laboratory System (UNOLS, http://www.unols.org). These vessels range in size, endurance, and capabilities, enabling NSF and other federally-funded scientists with the means to conduct ocean science research with a diverse fleet capable of operating in coastal and open ocean waters. UNOLS is an organization of 61 academic institutions and National Laboratories involved in oceanographic research and joined for the purpose of coordinating oceanographic ships' schedules and research facilities. The UNOLS charter was written in 1972. Each vessel is managed and operated by an academic institution in coordination with the UNOLS office at the University of Rhode Island. There are four classes of UNOLS vessels: global, ocean/intermediate, regional, and coastal/local.[113] The 6 global class vessels are large, high-endurance ships that work worldwide and can stay at sea for 50 days or more. These vessels carry 30-38 scientists on each expedition and range in length from 235-279'. The five ocean/intermediate class ships are medium-endurance vessels with berths for 18-25 scientists. Expeditions on these vessels, which are 168-186' in length, can last up to 40 days. There are 3 regional class vessels, which typically work closer to the coasts and are used for shorter cruises. These vessels are 135-146' in length. Finally, the six coastal/local class vessels are 66-125' in length and are used for short, near-shore cruises that are <20 days in length. Coastal/local vessels only have berthing for 15 or fewer scientists and are commonly used for educational purposes.

2. **EarthScope:** The EarthScope facility is a distributed, multi-purpose geophysical instrument array that is helping researchers make major advances in our knowledge and understanding of the structure and dynamics of the North American continent. EarthScope seeks to enhance our understanding of the structure and evolution of the North American continent, including earthquakes and seismic hazards, magmatic systems and volcanic hazards, lithospheric dynamics, regional tectonics, continental structure and evolution, fluids in the crust, and associated educational aspects. EarthScope is cyber-infrastructure enabled and is comprised of three parts: the USArray, the Plate Boundary Observatory, and the SAFOD drilling program. EarthScope began operations in 2009 and supports both fixed- and mobile sites and both PI-driven and "exploratory" science. EarthScope streams and archives data online, and a set of physical samples are available from SAFOD. The US Geological Survey (USGS), the National Aeronautics and Space Administration (NASA), the Department of Energy (DOE), and the International Continental Scientific Drilling Program are partners with NSF in EarthScope. Project partners also include state and local governments, geological and engineering firms, and

[112] These descriptions are modified from the FY2011 NSF Budget Request to Congress

[113] See the ships and their classes at http://www.unols.org/info/vessels.htm. Class descriptions are modified from the 2009 NRC report, "Science at Sea: Meeting Future Oceanographic Goals with a Robust Academic Research Fleet." UNOLS reclassified its ships in 2011.

Table B-1: Categorization of NSF's Operational Large Facilities

Facility	ARF	Arecibo	CHESS	EarthScope	IceCube	IRIS	IODP	Gemini	LIGO	LHC	NCAR	NEES	NHMFL	NNIN	NOAO	NRAO	NSCL	NSO	USAP
Year Began Operations	1971	1971[2]	1980[3]	FY 2009	2011	1984	2003	2000/2002	1992	2009	1960	2004	1993?	2004	1982	1956	1981	1962	
Predecessor/Pilot	historical vessels	other radio dishes	CESR	IRIS, UNAVCO, ICDP, geodetic networks	AMANDA	n/a	ODP	other scopes	multiple	CERN accelerator	indiv. Programs	scattered facilities	MIT Bitter Lab	National Nanofabrication Users Network	other scopes	other scopes	other cyclotrons	?	n/a
Facility Site Type											*FFRDC*				*FFRDC*	*FFRDC*			
Single Site																			
Distributed Single Sites																			
Network																			
CyberInfrastructure (CI)																			
CI-Enabled?	no[1]	no	no	yes	no	yes	no	no	no	no	no	yes	no	no	no	no	no	no	no
Facility Site Location																			
Domestic																			
International																			
Facility Site Mobility																			
Fixed																			
Mobile																			*
Research Disciplines																			
Archaeology																			
Astronomy																			
Physics																			
Chemistry																			
Materials Science																			
Engineering								?	?	?	?					?	?		?
Health Science																			
Atmospheric Sciences																			
Geology/Geochemistry																			
Oceanography																			
Climate Science																			
Polar Science																			
Biology																			
Ecology																			
Facility Use Model																			
Experiment																			
Service Facility																			
Network																			
On Line Data Archive?																			
Yes																?			
No		?							?	?									
Type of Use																			
On-Site																			
Remote Access		?				?								?	?				
Funding Sources/Partners																			
Interagency - NASA																			?
Interagency - DOE																			?
Interagency - NOAA																			
Interagency - NIH																			
Interagency - USGS																			
Interagency - DOD																			
Interagency - State																			
Industry/Private				?															
International Partner												?							
State/Local/Tribal							?												
Other																			

1. some vessels have stream visual data and enable remote participation via telepresence. Archived data available through R2R database
2. year became NSF's NAIC
3. additional detectors added between 1980-1999

Canadian and Mexican agencies. EarthScope is managed by the Incorporated Research Institutions for Seismology (IRIS) and UNAVCO.

3. **NEES:** The Network for Earthquake Engineering Simulation (NEES) is a national, networked simulation resource of 14 advanced, geographically distributed, multi-user earthquake engineering research experimental facilities with telepresence capabilities. NEES provides a national infrastructure to advance earthquake engineering research and education through collaborative and integrated experimentation, computation, theory, databases, and model-based simulation to improve the seismic design and performance of US civil infrastructure systems. Experimental facilities include shake tables, geotechnical centrifuges, a tsunami wave basin, large-scale laboratory experimentation systems, and mobile and permanently installed field equipment. NEES facilities are located at academic institutions (or at off-campus field sites) throughout the United States. NEES completed construction on September 30, 2004, and opened for user research and education projects on October 1, 2004. NEES is currently managed by NEESComm at Purdue University.

4. **NOAO/KPNO:** The National Optical Astronomy Observatory (NOAO) was established in 1982 by uniting operations of the Kitt Peak National Observatory (KPNO) in Arizona and the Cerro Tololo Inter-American Observatory (CTIO) in Chile. NOAO is a Federally Funded Research and Development Center (FFRDC) for research in ground-based, nighttime, optical, and infrared (O/IR) astronomy. NOAO also is the gateway for the US astronomical community to the International Gemini Observatory and to the "System" of federally-funded and non-federally-funded O/IR telescopes through the Telescope System Instrumentation Program (TSIP) and the Renewing Small Telescopes for Astronomical Research (ReSTAR) program. For all NOAO and "System" telescopes, peer-review telescope allocation committees provide merit-based telescope time but no financial support. NOAO telescopes are open to all astronomers regardless of institutional affiliation on the basis of peer reviewed observing proposals. NOAO is managed by the Association of Universities for Research in Astronomy (AURA). This study focuses on the KPNO branch of NOAO: KPNO began operations in the early 1960s and NSF supports three telescopes on KPNO (the Mayall 4-meter, the 2.1-meter, and the WIYN 3.5-meter telescopes).

Observed Meetings

During the course of this study, several meetings coordinated by or in conjunction with the Large Facilities Office provided useful insight into facility management, oversight, and use. These meetings are listed below (Table B-2).

Table B-2: Observed Meetings Relevant to Study

Facility	Meeting	Date
UNOLS	Business Systems Review: *R/V Langseth*	September 20, 2011
EarthScope	EarthScope Operations & Maintenance Review	September 27-28, 2011
UNOLS	UNOLS Fleet Improvement Meeting	October 24, 2011
UNOLS	UNOLS Annual Meeting	October 25-26, 2011
All	NSF Large Facility Recompetition Meeting	November 2-3, 2011
NEES	NEES MAST Lab NSF Site Visit	December 15-16, 2011
NEON	NEON Operations Review	January 3-6, 2012
All	National Science Board February Meeting	February 2-3, 2012
All	Annual Large Facilities Workshop	April 24-26, 2012
All	National User Facility Organization Annual Meeting	June 18-20, 2012

Site Visits and Interviews with Facility Staff
Several site visits to case study facilities were used to conduct interviews with facility staff (Table B-3). All visits were arranged independently with the exception of the visit to the NEES-MAST Lab, where interviews were conducted after a routine NSF site visit organized by the program manager. Interviewees represent a sampling of facility staff (Table B-4). Interviews ranged from approximately 30 to 60 minutes in length.

Table B-3: Facility Site Visits and Meetings

Facility	Location	Date of Visit
EarthScope/IRIS	IRIS Headquarters, Washington, DC	November 16, 2011
NEES-MAST Lab	Univ. of Minnesota, Minneapolis, MN	December 16, 2011
EarthScope/UNAVCO	UNAVCO, Boulder, CO	January 6-10, 2012
UNOLS	AGU Ocean Sciences Meeting, Salt Lake City, UT	February 20-24, 2012
NEESComm	Purdue University, West Lafayette, IN	March 7-9, 2012
NOAO/KPNO	Univ. of Arizona/Kitt Peak, Tucson, AZ	March 14-16, 2012
STScI	Johns Hopkins Univ., Baltimore, MD	May 17, 2012
DOE	DOE Basic Energy Sciences, Germantown, MD	May 30, 2012

Interview Questions
During each interview, a series of questions focused on the interviewee's role within the facility and facility utilization were discussed. Questions varied depending on the position and responsibilities of the interviewee. A sample set of questions is below.

1. What is your role with the facility?
2. What is your definition of a facility user?
3. Who are the facility's users (specific examples?)?
4. Can you provide some examples of how users provide feedback (both formally and informally)? (e.g., user groups, talking w/community at professional meetings, online forums, etc)?
5. Do you track users? If so, how? Why? Is this important to the facility? If so, how is this information used?
6. Have the users for which the facility was built used the facility? Have there been any user groups you expected to use the facility that did ultimately not participate?
7. Have there been any unanticipated users? (specific examples)?
8. Does the facility try to recruit new users? How is this done? How do you make decisions about including new users?
9. What are some mechanisms in place for training new users?
10. Do you consider educators to be users? Do you consider the public to be users? Do you consider technical staff to be users?
11. How has use changed over time in your experience using/working with the facility?

Analyzing UNOLS User Data

For this study, the UNOLS office provided the author with 10 years of data collected from their cruise reports. UNOLS asks each ship operator to complete one of these reports after each expedition and questions include basic information such as expedition dates and ports as well as user information, such as the primary disciplines of the research conducted on board to the numbers of shipboard crew members organized by their primary function (scientist, graduate student, undergraduate, technicians, observer, etc.). UNOLS is the recipient and keeper of these data. However, not all operators complete a report for every expedition: for example, in 2010, a total of 16 ships out of 21 submitted reports. Therefore, the available data is incomplete and percentages are used to account for data discrepancies. Here, the percent "user" (as defined by UNOLS functions listed in the survey, such as scientist, graduate student, undergraduate, etc.) was calculated for each year using the data available. Data was cleaned to calculate the percentages by removing unresolvable data (e.g., number reported as “1/2” rather than 1, 2, or 0.5). Detailed steps of these analyses and accompanying spreadsheets are available by contacting the author.

Table B-4: Interviewees

#	Facility	Title	Institution	Date	Location
1	All	Assistant Director, Physical Sciences	OSTP	02/09/2012	Washington, DC
2	All	Senior Policy Analyst	OSTP	02/09/2012	Washington, DC
3	All	Assistant Director for Federal Research and Development	OSTP	02/09/2012	Washington, DC
4	All	Program Manager	NSF	10/28/2011	Arlington, VA
5	All	STPI Staff	STPI	11/30/2011	Washington, DC
6	ALMA	Program Manager	NSF	Oct, 2011	Arlington, Va
7	DOE	Senior Technical Advisor, Basic Energy Sciences	DOE	05/30/2012	Germantown, MD
8	DOE	Senior Science and Technology Advisor, Office of Science	DOE	05/30/2012	Germantown, MD
9	DOE	Director, Basic Energy Sciences	DOE	05/30/2012	Germantown, MD
10	DOE	Prog. Mgr, Facility Coordination, Metrics, Assessment, BES	DOE	05/30/2012	Germantown, MD
11	EarthScope	Program Manager	NSF	11/01/2011	Arlington, Va
12	EarthScope	Researcher	USGS	01/10/2012	Lakewood, CO
13	EarthScope	Project Manager III	UNAVCO	01/09/2012	Boulder, CO
14	EarthScope	Data Manager	UNAVCO	01/09/2012	Boulder, CO
15	EarthScope	Data Products Manager	UNAVCO	01/09/2012	Boulder, CO
16	EarthScope	Director of International Development Seismology	IRIS	11/27/2011	Washington, DC
17	EarthScope	GPS Operations Manager	UNAVCO	01/09/2012	Boulder, CO
18	EarthScope	Staff Scientist/Chair of USArray User Committee	Carnegie	11/18/2011	Washington, DC
19	EarthScope	PBO Director	UNAVCO	01/06/2012	Boulder, CO
20	EarthScope	Project Manager III	UNAVCO	01/09/2012	Boulder, CO
21	EarthScope	President	UNAVCO	01/09/2012	Boulder, CO
22	EarthScope	E&O Specialist	UNAVCO	01/09/2012	Boulder, CO
23	EarthScope	President	IRIS	11/27/2011	Washington, DC
24	EarthScope	Cost Schedule Manager	UNAVCO	01/09/2012	Boulder, CO
25	EarthScope	Education and Outreach Program Manager	IRIS	11/27/2011	Washington, DC
26	EarthScope	USArray Director	IRIS	11/27/2011	Washington, DC
27	HST	Library Technician/Bibliographer	STScI	05/17/2012	Baltimore, MD
28	HST	Product Development, Virtual Astronomy Observatory	STScI	05/17/2012	Baltimore, MD
29	HST	Chief Librarian	STScI	05/17/2012	Baltimore, MD
30	HST	Mission Head, Community Missions Office	STScI	05/17/2012	Baltimore, MD
31	HST	Scientist	STScI	05/17/2012	Baltimore, MD
32	HST	Mission Head, HST Mission Office	STScI	05/17/2012	Baltimore, MD
33	HST	MAST Manager	STScI	05/17/2012	Baltimore, MD
34	LHC	Program Manager	NSF	10/24/2011	Arlington, Va
35	LHC	Program Manager/Education Point of Contact	NSF	10/24/2011	Arlington, Va
36	NEES	Director, Education, Outreach, & Training	Purdue University	03/08/2012	West Lafayette, IN
37	NEES	Research Fellow	Univ. of Minnesota	12/16/2011	Minneapolis, MN
38	NEES	IT Manager	Univ. of Minnesota	12/16/2011	Minneapolis, MN
39	NEES	Researcher, User Committee Chair	Purdue University	03/09/2012	West Lafayette, IN
40	NEES	IT Manager and Co-PI	Univ. of Minnesota	12/16/2011	Minneapolis, MN
41	NEES	Deputy Director, NEESComm	Purdue University	03/08/2012	West Lafayette, IN
42	NEES	Co-PI and EOT Manager	Univ. of Minnesota	12/16/2011	Minneapolis, MN
43	NEES	Research Coordinator	Univ. of Minnesota	12/16/2011	Minneapolis, MN
44	NEES	Co-Leaders for IT, NEESComm	Purdue University	03/09/2012	West Lafayette, IN
45	NEES	Director, Site Operations	Purdue University	03/08/2012	West Lafayette, IN
46	NEES	Site Operations Engineer	Purdue University	03/08/2012	West Lafayette, IN
47	NEES	Program Manager	NSF	11/21/2011	Arlington, Va
48	NEES	PI and Lab Manager	Univ. of Minnesota	12/16/2011	Minneapolis, MN
49	NEES	Project Manager	Univ. of Minnesota	12/16/2011	Minneapolis, MN
50	NEES	User Committee Member/Site Operations Engineer	Purdue University	03/08/2012	West Lafayette, IN

51	NEES	Director, IT Management	Purdue University	03/09/2012	West Lafayette, IN
52	NEON	Project Manager	NEON, Inc.	01/06/2012	Boulder, CO
53	NEON	Chief Science Officer	NEON, Inc.	01/03/2012	Boulder, CO
54	NOAO	Program Manager	NSF	02/09/2012	Arlington, Va
55	NOAO/KPNO	Deputy Director	NOAO/KPNO	03/15/2012	Tucson, AZ
56	NOAO/KPNO	Telescope Operations Manager	NOAO/KPNO	03/16/2012	Tucson, AZ
57	NOAO/KPNO	Director	NOAO/KPNO	03/16/2012	Tucson, AZ
58	NOAO/KPNO	Observing Asst (Telescope Operator)	NOAO/KPNO	03/15/2012	Tucson, AZ
59	NOAO/KPNO	Manager, Facilities	NOAO/KPNO	03/14/2012	Tucson, AZ
60	NOAO/KPNO	Manager, Visitor's Center	NOAO/KPNO	03/14/2012	Tucson, AZ
61	NOAO/KPNO	Engineer	NOAO/KPNO	03/16/2012	Tucson, AZ
62	NOAO/KPNO	Astronomer	NOAO/KPNO	03/16/2012	Tucson, AZ
63	NOAO/KPNO	Scientist/Instrument Development	NOAO/KPNO	03/15/2012	Tucson, AZ
64	NOAO/KPNO	Head, Education & Public Outreach	NOAO/KPNO	03/14/2012	Tucson, AZ
65	NOAO/KPNO	Observing Asst (Telescope Operator)	NOAO/KPNO	03/15/2012	Tucson, AZ
66	NOAO/KPNO	Senior Scientific Programmer	NOAO/KPNO	03/14/2012	Tucson, AZ
67	NOAO/KPNO	Researcher	NOAO/KPNO	03/15/2012	Tucson, AZ
68	NOAO/KPNO	Public Programs Specialist	NOAO/KPNO	03/15/2012	Tucson, AZ
69	NOAO/KPNO	Researcher	NOAO/KPNO	03/15/2012	Tucson, AZ
70	NOAO/KPNO	Observing Superintendant	NOAO/KPNO	03/15/2012	Tucson, AZ
71	NSO	Program Manager	NSF	Oct., 2011	Arlington, Va
72	NUFO	Steering Committee Member	NUFO	03/26/2012	Washington, DC
73	NUFO	Chair	NUFO	03/26/2012	Washington, DC
74	OOI	CI Manager, OOI	SIO	02/22/2012	Salt Lake City, UT
75	UNOLS	Executive Secretary	UNOLS	11/30/2012	Washington, DC
76	UNOLS	Data Manager, R2R	LDEO	02/23/2012	Salt Lake City, UT
77	UNOLS	Assistant Executive Secretary	UNOLS		by email
78	UNOLS	Program Manager	NSF		Arlington, VA

APPENDIX C: ACRONYMS

AAAS	American Association for the Advancement of Science	**NEON**	National Ecological Observatory Network
AAS	American Astronomical Society	**NNIN**	National Nanotechnology Infrastructure Network
AGU	American Geophysical Union	**NOAO**	National Optical Astronomy Observatory
AIBS	American Institute of Biological Sciences	**NRC**	National Research Council
ALMA	Atacama Large Millimeter Array	**NSB**	National Science Board
ARF	Academic Research Fleet	**NSF**	National Science Foundation
ARM	Atmospheric Radiation Measurement Climate Research Facility	**NUFO**	National User Facilities Organization
BER	Basic Energy Research (DOE)	**O/IR**	Optical/Infrared Wavelengths
BES	Basic Energy Sciences (DOE)	**OMB**	White House Office of Management & Budget
CI	Cyberinfrastructure	**OOI**	Ocean Observatories Initiative
DOE	Department of Energy	**OSTP**	White House Office of Science & Technology Policy
ESA	Ecological Society of America	**PBO**	Plate Boundary Observatory
FFRDC	Federally Funded Research and Development Center	**R&RA**	Research and Related Activities
HST	Hubble Space Telescope	**R2R**	Rolling Deck to Repository Database
IODP	Integrated Ocean Drilling Program	**SAFOD**	San Andreas Fault Observatory at Depth
IP	Internet Protocol	**STPI**	Science & Technology Policy Institute
IRIS	Incorporated Research Institutions for Seismology	**STScI**	Space Telescope Science Institute
KPNO	Kitt Peak National Observatory (part of NOAO)	**UEC**	User Executive Committee
MAST	Multi-Axial Subassemblage Testing Laboratory	**UNOLS**	University National Oceanographic Laboratory System
MREFC	Major Research Equipment and Facilities Construction	**USAP**	US Antarctic Program
NASA	National Aeronautics and Space Administration	**VAO**	Virtual Astronomy Observatory
NEES	George E. Brown Network for Earthquake Engineering Simulation	**VO**	Virtual Observatory