Report to the National Science Foundation

from the

Sixth Workshop on the Development of a National Ecological Observatory Network (NEON): Information Management

held at the

National Center for Ecological Analysis and Synthesis,
University of California Santa Barbara,
Santa Barbara, CA
September 16-18, 2002

I. Executive Summary

The overarching goal of NEON is to provide, for the first time, a highly coordinated national system for simultaneously monitoring a number of critical ecological and environmental measures at fine grain over large spatial and temporal scales. NEON data will enable new types of ecological and environmental science by providing researchers and resource managers with access to exceptionally high quality information that is comprehensive, timely, and accurate. NEON is envisioned to be an **integrated platform** for observing the environment in support of the broadest range of scientific studies. The NEON framework will lead to fundamentally new scientific insights; more robust, predictive science; and vastly enhanced accuracy in forecasting environmental trends. The network will be available to a broad array of scientists who seek opportunities for cross-disciplinary research. Consequently, one of the key features of NEON will be information systems that promote the broadest levels of data exchange and knowledge development in the environmental research community.

To meet these goals, the functional characteristics of NEON information management necessarily include support for data acquisition, quality assurance and quality control, storage and archiving, data dissemination and access, data integration and aggregation, analysis, synthesis, and modeling. These are ambitious goals for an information management system, especially one dealing with highly dispersed and heterogeneous data. To realistically meet them, the investment in NEON Information Management will need to be substantial: for each observatory, 25-40% of the annual budget and 50-75% of the budget for the NEON Coordinating Unit. Although this is a substantial portion of the funds supporting NEON, the program will not be a network without comprehensive information management.

Enabling NEON observatories to function as an integrated network will require extensive coordination among the observatories engendered by extensive network-wide planning. Consequently, the workshop participants strongly endorsed the need for a NEON Coordinating Unit that could fill such a role. The need for a Coordinating Unit is particularly obvious for information management, as the flow of data across the observatories and, perhaps more importantly, to the broader research community outside of NEON, is a crucial way in which NEON would function as a network.

The information management roles for the Coordinating Unit should include data archiving, standards development and promulgation, development of a national NEON portal for accessing NEON resources, knowledge engineering, network design, operation, and performance, and high performance computing. **Participants believed strongly that the NEON Coordinating Unit should be created prior to, or at least no later than, the first NEON observatories, and that carefully crafting the responsibilities, authority, and accountability of the coordinating unit was among the most critical tasks in making NEON successful.** The participants raised many questions about the design of the most effective governance structure for lines of authority within NEON. Participants discussed the need for reviews of network effectiveness and participation on a frequent (semi-annual) basis, rather than the typical site review model that spans several years in other NSF networks. More frequent, internal reviews in which there is a mechanism for enforcement authority by the review body was considered essential to developing an integrated network rather than a loosely federated set of nodes.

Participants agreed that centralized development of an information management infrastructure at the Coordinating Unit would be far more efficient and effective than distributed development. However, participants raised several important issues about the need for direct involvement of the observatories in the design and implementation of network-wide infrastructure, and the need for observatories to be able to explore areas where they have unique needs.

A fundamental goal for NEON is to provide timely and broad access to all data. High quality data, permanent and persistent archives, rapidly available and highly usable data stores, interoperability, standardization, and accountability should be hallmarks of the NEON scientific enterprise. All data collected under the auspices of NEON shall be documented and archived on NEON information systems. Data from core NEON measurements must be openly available and accessible to all.

Data systems need to accommodate the breadth and heterogeneity of data found across the environmental sciences, from molecular genetics to ecosystem data. This implies a strong reliance on standardized data acquisition protocols and metadata descriptions of data, and flexible software for analysis and management that are metadata-driven. Participants believe strongly that there should be an emphasis on standards for information management, and that observatories should use a single shared infrastructure that is developed to meet the needs of all of the observatories. Participants recognized that this generic architecture would need to accommodate links to consortium institutions that may be providing resources through one of the observatories and have existing information management systems. Consequently, the interfaces of NEON information systems should be standardized first, and secondarily implementations of these standardized interfaces should be developed.

The network infrastructure for NEON will need very high bandwidth at main observatory sites and some satellite sites, depending on their role. All field sites at observatories will need tiered wireless networks for sensor deployment and support of interactive applications for scientists. NEON needs to use a network topology that supports high-throughput, fail-safe data transfer among observatories, the coordinating unit, and the broader research community.

Remotely sensed images must be an integral component of the core monitoring program for a NEON observatory. The network should possess a high level of expertise and be able to contribute significantly to advances in the management and interpretation of image data. In addition, the accurate geo-referencing of all field data collected through a NEON facility is essential to the ability to understand the spatial context of ecosystem function and the scalability of these observations to broad, regional questions.

The provision of advanced and standardized approaches to computation, analysis, and synthesis is one of the key advantages of NEON, and will have a major beneficial impact for NEON-affiliated researchers, as well as the general ecological community. NEON should adopt emerging GRID-based approaches towards unifying the computing environment, provide high-performance compute servers on which to run analyses and models, and train researchers in high-performance computing techniques. Finally, NEON should also develop standard software interfaces for visiting or non-aligned systems to access NEON services.

<u>Cite as</u>:
Jones , M.B., B. Benson, P. Bryant, P. McCartney, W. Michener, and M. Schildhauer. 2003. Report to the National Science Foundation from the Sixth Workshop on the Development of a National Ecological Observatory Network (NEON): Information Management.

II. Table of Contents

I. Executive Summary i
II. Table of Contents iii
III. Introduction 1
IV. NEON Coordinating Unit 2
Role of the NEON Coordinating Unit 2
Authority, Accountability, and Governance 3
V. Data Policies 4
VI. Data Acquisition and Quality Management 4
Distribution of functions 5
Enforcement of standards 5
Tool development 5
Other issues 5
VII. Databases, Metadata, and Archiving 6
Hardware and Software needs 8
Knowledge Engineering, Semantics and Integration 8
VIII. Networking and Sensors 8
Networking within Observatories 9
Wireless networking 9
Networking satellite sites 9
Telemetry & sensor networks 10
Networking among Observatories 11
Networking to the Broader Research Community 11
IX. Remote Sensing and Geographic Information Systems 12
Data acquisition 12
Processing 12
Storage and management 13
Research integration 13
Priorities for Archiving 13
X. Computation, Analysis, and Synthesis 13
General Computational Needs 14
Distribution of responsibilities 15
Consultation and Training 16
Specific Computational Needs 17
Observatory 17
Documentation 17
Software 17
XI. Personnel 18
Observatory Personnel 18
Coordinating Unit Personnel 18
XII. Summary 19
XIII. Appendix: Workshop Participants 20

III. Introduction

Like previous workshops, the participants enthusiastically endorsed the concept of NEON. The concept of NEON as an integrated research platform providing a national scale source of information on all aspects of biological information characterizing our environment is exciting in terms of the new science that it will enable, but also challenging in terms of information management. Participants at this sixth workshop were tasked with evaluating the information management needs of the NEON at both the network level and the observatory level in order to accommodate the broad range of information (molecular, ecological, physical, chemical, etc.) that would be collected in NEON.

Participants noted that NEON is different from existing networks such as the LTER in that the funds for research goals would be separated from funds for creating and maintaining the observatory network. In short, NEON is envisioned to be an **integrated platform** for observing the environment, and should support the broadest range of scientific studies. This represents a fundamental difference in the way environmental science is funded, in that observatories cannot be evaluated independent of the NEON as a whole.

Given that the overarching goals of NEON are to provide data of high utility for addressing critical ecological and environmental issues, like the nature and pace of biological change, the information management goals for NEON are:

- **end-to-end quality** — support collection and maintenance of high quality data and accompanying metadata from the field to the laboratory and through analysis and dissemination
- **timeliness** — provide timely and broad community access to NEON data and information (open and real-time data access where appropriate)
- **relevance** — facilitate knowledge transfer to scientists, decision-makers, and the public (e.g., provide customized views to make data and information, such as summaries, trends, graphics, and visualizations, accessible to these stakeholders)
- **technology leadership** — facilitate NEON network coordination and collaboration (i.e, instrumentation, data collection, and analytical protocols; wireless and automated data collection and processing)
- **interoperability** — provide a tightly coupled national network; interoperability is paramount
- **standardization** — design and develop <u>or</u> identify and adopt data and information standards for core measurements
- **flexibility** — build information management systems that are robust and extensible—which may be best designed through modularity and well designed interfaces among components
- **discovery** — facilitate and enhance the data discovery process
- **security** — provide for secure and long-term access to the data (backup, mirroring)
- **stewardship** — create a national data resource that can be mined for decades to come
- **collaboration** — create an environment for data quality that draws upon collaboration among scientists, technicians, information technologists, and all personnel engaged in the NEON enterprise
- **research leadership** — promote collaborative research into relevant information technologies that meet NEON objectives

To meet these NEON goals, the functional characteristics of NEON Information Management necessarily include support for:

- data acquisition—acquiring data through sensors, field measurements, etc.
- quality assurance and quality control
- storage and archiving
- dissemination and access to data, information, and knowledge
- integration and aggregation
- analysis, synthesis, and modeling.

These are ambitious goals for information management system, and to realistically meet them, the investment in NEON Information Management will need to be substantial:

- each observatory: 25-40% of annual budget should be allocated for a functional information management system
- NEON Coordinating Unit: 50-75% of annual budget

IV. NEON Coordinating Unit

Enabling the collection of NEON observatories to truly function as an integrated network will certainly require extensive coordination among the observatories and extensive network-wide planning. Consequently, the workshop participants strongly endorsed the need for a NEON Coordinating Unit that could fill such a role. The need for a coordinating unit is particularly obvious for information management, as the flow of data across the observatories and, perhaps more importantly to the broader research community outside of NEON, is a crucial way in which NEON would function as a network.

Role of the NEON Coordinating Unit

The roles for a NEON coordinating unit that were discussed include the following:

1. Administration, including accounting, management, and staffing
2. Network-wide Purchasing Agent (for economies of scale), specifically targeting large equipment and remote sensing data for cost savings
3. Meeting Coordination for the NEON All Scientist's Meetings, Science and Technology Advisory Boards, and Standing Committees
4. Training Programs including courses, manuals, and documentation on subjects such as instrumentation, informatics, analytical approaches, modeling, molecular techniques, meteorological measurements, and analytical and isotopic chemistry. Some of these could be distance learning programs
5. Informatics Services
   a. data archive center (persistent archives, mirroring, reliability, standards, data formats)
   b. standards development and promulgation (e.g., QA/QC standards, metrics, tools, QA/QC plans published, QA/QC consulting and training)
   c. standard methods for archiving and curation of samples, data, and information
   d. National NEON portal for accessing NEON resources
   e. NEON knowledge engineering—adding value to NEON data, providing links to other relevant data and information
   f. Standards for data representation and abstraction to facilitate network discovery and data-reuse
   g. Communications, including network operation and performance
   h. Facilitate high performance computing and communication
   i. Cross-site data preparation
6. Public Relations (publicity, newsletter, web content) in order to:

   a. accommodate and encourage participation by the broad research community
   b. facilitate public outreach and education
7. Network Coordination (i.e., facilitate use of NEON by the broader multidisciplinary community and ensure that the NEON information base is supportive of national and continental scale research). Provide a common framework for collaboration. Coordination activities include:
   a. coordination with existing programs and organizations
   b. establish, implement, and evaluate standards for bioinformatics, core measurements, and equipment
   c. review new developments in science and technology
   d. coordinate information technology and management
   e. coordinate efforts with international bodies like GBIF and IABIN
   f. link personnel and capabilities among sites
   g. evaluate efforts in research and education
   h. maintain common framework for network collaboration and management
   i. evaluate network success in terms of data gains, science gains, technology gains, educations gains, and process gains
   j. Organize standing committees (users, technology, data, education)

Authority, Accountability, and Governance

Although the need for a coordinating unit in the NEON was very clear to participants, a number of difficult issues regarding authority and accountability of the coordinating unit were discussed. There was general agreement that, in order for NEON to fulfill its potential, **the NEON Coordinating Unit should be created prior to, or at least no later than, the first NEON observatories**. Participants felt that this timing was critical because the Coordinating Center will be at a disadvantage in trying to set network-wide standards if a collection of observatories are already in existence with their own established procedures before the Coordinating Unit is funded. Grafting a network concept onto a group of previously established sites will not automatically create a network effect. **Workshop participants felt that carefully crafting the responsibilities, authority, and accountability of the coordinating unit was among the most important tasks in making NEON successful**.

Some of the issues to be resolved include:

- How much authority does the NEON Coordinating Unit have in terms of evaluating observatories' participation in the network, and enforcing network-wide standards in informatics and data collection methods?
- What mechanism(s) can be used to ensure that the NEON Coordinating Unit is accountable to the observatories and effectively uses its resources for the benefit of the whole network?
- There was broad consensus that full participation by NEON observatories and the Coordinating Unit would be required for effective and optimal network design, governance, and scientific agendas. However, there were questions of what would be the most effective governance structure for lines of authority within NEON. Participants discussed the need for *internal* reviews of network effectiveness and participation on a frequent (semi-annual) basis, rather than the typical external site review model in other NSF networks that spans several years. More frequent, *internal* reviews in

which there is a mechanism for enforcement by the internal reviewers was considered essential to developing an integrated network rather than a loosely federated set of nodes.

- There was consensus that centralized development of an information management infrastructure would be far more efficient and effective than distributed development. However, participants raised several important issues about the need for tight involvement of the observatories in the design and implementation of network-wide infrastructure, and about the need for observatories to be able to explore areas where they have unique needs.

V. Data Policies

Two types of data are collected in the NEON enterprise: core measurements collected throughout the network and experimental data collected by individual or groups of scientists that visit observatories. **The overarching goal for NEON is to provide timely and broad access to all data**. High quality data, permanent and persistent archives, rapidly available and highly usable data stores, interoperability, standardization, and accountability are hallmarks of the NEON scientific enterprise. All data collected under the auspices of NEON shall be documented and archived on NEON information systems. Data from core NEON measurements are openly available and accessible to all. Furthermore, standard operating procedures shall be developed for collection and quality assurance of core measurement data. Data from other research conducted under the auspices of NEON will be made available in a timely fashion with requisite metadata. For example, (1) data collected under the auspices of a project funded in association with a NEON are to be made available prior to the termination of the project; (2) data underpinning any scientific publication are to be made available at the time of publication.

NEON data policies should promulgate a cultural change that values data sharing. NEON user policies and agreements should require appropriate acknowledgment and attribution for data providers and funding sources (e.g., NSF and NEON); policy should govern fair use of data by data users.

A data policy committee of scientists and technologists shall be comprised to: (1) approve exclusions (e.g., data dealing with human subjects, locations of endangered species); (2) develop appropriate data use agreements (e.g., data citation and acknowledgement of data providers and funding support); and (3) develop appropriate information protocols for sample and voucher specimens.

Network-wide data and metadata standards shall be developed and adopted for use by NEON scientists. Emphasis shall be placed on adopting or developing community standards, enhancing the ability for machine parsing of data and metadata, and accommodating well-defined legacy standards. Standards and policies (including enforcement actions) shall be developed by the NEON Coordinating Unit in conjunction with NEON observatories (and NSF as appropriate). The governance structure for NEON observatories and the NEON Coordinating Unit shall be designed to promote and enforce adherence to data sharing and information management policies.

VI. Data Acquisition and Quality Management

Preeminently high quality data should be one of major strengths of the NEON program. The development of quality assurance/quality control (QA/QC) standards for core data and the adherence to those standards as well as the development of tools to facilitate implementation should be a top priority for the initial NEON observatories, and initial budgets must allocate resources to this endeavor.

Distribution of functions

Responsibilities for quality management reside at the Coordinating Unit, at each NEON observatory, and with the researchers who will be conducting research at NEON observatories. The Coordinating Unit will provide QA/QC standards and metrics to be applied across the NEON network. The informatics staff at the Coordinating Unit will develop tools to be used to support the high data quality in areas such as metadata capture and automated QA/QC checking. Each NEON site is responsible for publishing a QA/QC plan (e.g., as an appendix to their proposal and/or on their website) that meets the NEON standards. The site will also provide support staff to provide training in QA/QC tools and standards.

It is necessary to set expectations for researchers who will be using the NEON site yet not burden the researchers with unreasonable requirements that discourage the use of the observatory. Each NEON observatory will need to establish rules for doing research at the observatory. Metadata should be required for each research project. Experience shows that acquiring most of these metadata through a process of project registration at the beginning of the research is most effective. The metadata should include QA/QC information. Another observatory use rule should be that each project must be geo-referenced. Beyond rules such as these, high quality data from researchers can be facilitated by making QA/QC tools, expertise, and training available. IT staff at the NEON observatory will explain policies, tools, instruments, and metadata requirements.

Enforcement of standards

The primary mechanism for enforcement of standards will be to make examination of QA/QC a formal part of the observatory and NSF review process, thus making funding contingent on adherence to standards. Other possible mechanisms include review by additional independent third parties and generating QA "report cards" that assess performance. The need for faster review cycles also became apparent for QA/QC; therefore, NEON should develop mechanisms for short-term review and accountability in order to facilitate NEON's objectives.

Tool development

The requirement for QA/QC standards across the NEON observatories and the amount of core data to be subjected to QA/QC necessitates the development of tools that maintain and automate the process. Data streaming in real-time from spatially distributed sensors should be automatically screened for data quality and equipment failures and processed through data cleansing algorithms. Development of such software in the early years of NEON could benefit from establishing relationships with computer scientists involved in areas such as anomaly detection working in partnership with the domain-based scientists with knowledge of the quality checking requirements/algorithms. Sensor performance can be modeled against previous output.

Generic software should be developed for researchers in the field using laptops or handheld devices that provides QA such as domain checks in real time. At a core NEON observatory, field researchers should have access to bi-directional wireless networks. With remote wireless connection to a server housing the database, more sophisticated QA checks against previously collected data are possible.

Other issues

Tracking versions of data sets is necessary after initial data release. Publication of a data set through a central repository requires parameters that describe status and access control. Annotations in metadata should contain update and review information as well as procedural changes, instrument changes, and calibration history.

Data recorded directly by people in the field can be subject to systematic bias. This bias may be hard to detect without a protocol for resampling, which should be considered in the design of the core measurements.

Standards for metadata and for the capture and sharing of video data (both still and motion) do not currently exist. The NEON network needs to establish these standards for the network.

Maintenance of sensors is essential to high quality data and requires scheduled preventive maintenance, calibration, and re-swapping of instruments. Redundant sensors can provide important QA information and protect against data loss.

Databases should include information on data quality for each datum and sensor. It is desirable where possible for quality information to be machine understandable, for example, through data flagging.

Finally, the NEON Coordinating Unit should provide tools that allow scientists to capture domain-specific quality assurance rules about their specific data. The rules should be formally expressed and used by automated processors to validate data collected either after the collection event or possibly in real-time as data are collected.

VII. Databases, Metadata, and Archiving

To achieve the NEON vision of a fully integrated network that supports seamless access to data by the broadest research communities, a highly coordinated information management strategy for data storage, data discovery, access control, and long-term preservation will be required. As stated in the data policy, both core data measurements and data produced by visiting researchers must be housed in and handled by the storage system and processing systems. These data (core and investigator-driven) combined represent an incredibly heterogeneous set of data to be managed that will evolve over time. The data from core measurements will include sensor-derived physical and chemical measurements associated with climate, hydrology, and gas fluxes, biodiversity monitoring data, ecological dynamics data, and genetic and other molecular level data. This breadth and heterogeneity will necessitate a flexible system that can be easily extended to allow for new types of data as researchers work at NEON over time. Participants agreed that such a system must have metadata at its core in order to enable both extensibility to novel data types and usability. This emphasis on metadata should extend throughout the scientific process, and include NEON-wide standards for data documentation, formal documentation for analytical processes and modeling, and formal documentation for methods and protocols.

Participants felt strongly that there should be a strong emphasis on standards for information management, and that, wherever possible, observatories should use a single shared infrastructure that is developed to meet the needs of all of the observatories. Participants recognized that this generic architecture would need to accommodate links to consortium institutions that may be providing resources through one of the observatories and have existing information management systems. Consequently, the interfaces of NEON information systems should be standardized first, and secondarily standard implementations of these interfaces should be developed. Standardized interfaces will tremendously simplify the process of linking existing systems to the NEON architecture without compromising the

NEON's ability to develop a holistic, integrated system. For example, metadata systems interoperability can be achieved by agreeing first on a communications protocol for exchanging metadata, second by agreeing on a syntax for the metadata exchange, and third by agreeing on the content of the metadata. This approach allows new metadata content to be introduced as the need arises without changes in the systems for exchanging and representing metadata. We also note that standardizing on particular software programs or hardware systems will not necessarily contribute to overall system interoperability. Instead, a focus on the data is required.

Workshop participants supported a model in which a data storage infrastructure that supports generic ecological data is developed for the NEON network that would be mandatory for each observatory and that unites the observatories into a seamless, cohesive network. This infrastructure may standardize the interfaces between data systems, and it may also standardize specific implementations where warranted. Design of this generic infrastructure would involve personnel from the Coordinating Unit as well as extensive, ongoing participation of the observatories through an information management committee. Development of the resulting infrastructure would occur primarily through a team of software engineers located at the Coordinating Unit. Deployment of this infrastructure would occur at the observatories through a systems integrator/developer who would be able to interface the generic management system to existing infrastructure at the observatory core and satellites. Again, we emphasize that the first layer of standardization would be at the application programming interface, and secondarily there would be standardized implementations in cases where that would result in substantial implementation efficiency gains.

There was substantial discussion of the extent to which a centrally developed solution could be expected to meet the needs of the observatories, and recognition that centralized, top-down models have failed in some particular situations in the past. Thus, it was deemed critical that the observatories be active participants in all aspects of system design and system evaluation. The clustering of development activities at the Coordinating Unit was intended to 1) create a critical mass of personnel to improve the development team's efficiency; 2) eliminate arbitrary redundancy in development efforts for core information management systems; and, 3) create a unified, seamless NEON-wide system in which interoperability among observatories is paramount.

Over the short term, the NEON system should focus on the storage, query, dissemination, and access control for heterogeneous data that includes the following features:

- **standardized metadata** for data, analyses, models, and methods, which should build upon the successful community of scientists that are starting to use Ecological Metadata Language
- **strong versioning** for data releases, including accession numbers (c.f., GenBank) and mechanisms for managing ongoing data sets within a strong versioning system
- **formal lineage tracking** of relationships between raw data products, derived data products, analysis and modeling code, etc.
- **long term preservation** of data for future, unforeseen syntheses and analyses, which implies extensive, high-quality, structured metadata and open data formats
- **near real-time access** to some data, tiered by researcher-set priorities
- **NEON-wide, shared authentication** and access control systems that leverage existing GRID efforts

- **formalized exchange protocols** for interfacing with existing infrastructure at observatories
- **scalable storage, query, and network systems** to support delivery of large data streams to the broadest research community (including non-NEON researchers)
- **standardized services** for data system interoperability

Hardware and Software needs

Participants recognized that there would be tremendous overlap in the computing needs of the NEON observatories because each will be collecting the full range of ecologically relevant data. Consequently, huge cost savings could be achieved by coordinating purchases of hardware and software across the network. The Coordinating Unit should play a major role in achieving these economies of scale.

Knowledge Engineering, Semantics and Integration

Even within a tightly coordinated network such as NEON, one would expect extreme heterogeneity in the information collected within and among sites. NEON observatories will be tasked with collecting data on all aspects of the environment across multiple levels of organization (microbial through ecosystem). The types of data collected would include genetic data from a variety of organisms, distribution and abundance of species, physical and chemical properties of the environment, behavioral information about organisms, and much more. These data will necessarily apply across habitat boundaries and environmental gradients, and will be collected at multiple temporal and spatial scales. Data will be collected in the context of monitoring, and within experimental regimes where particular aspects of the environment are manipulated, rendering the data appropriate for some types of analysis but not others.

Managing this extensive heterogeneity will be the foremost challenge for NEON information systems. Although existing networks such as the Knowledge Network for Biocomplexity (http://knb.ecoinformatics.org) provide a means for storing and documenting heterogeneous metadata and data by using Ecological Metadata Language (EML), the semantic issues associated with understanding data well enough for integration have not been effectively addressed. For example, in trying to use NEON, researchers will need to be able to quickly discover, integrate, and analyze data seamlessly from sources across national scales. This will necessitate a deep understanding of the data measurements, the protocols used in collecting the data, and the analytical constraints that particular analyses might impose. This information must be accessible to automated processors to be effectively used at national scales; currently it is available only directly from researchers, and partly through natural language descriptions of the data. For NEON to truly be an integrated network, this must change.

So, in addition to providing a rich technical infrastructure for NEON, we see an opportunity and a need for NEON to address the knowledge infrastructure of the network. This is an area where leadership and vision in information technology will be critical because computational solutions to many of the problems are not available. Nevertheless, the knowledge infrastructure of NEON will be the fundamental axis upon which researchers will judge the effectiveness of the network.

VIII. Networking and Sensors

NEON will be a network in several distinct ways. At its most fundamental it will need to be a physical network connecting sensors, data storage systems, and data processing systems within observatories, connecting these systems among the observatories that participate in NEON, and connecting to the broader scientific community via the Internet. The technologies used to connect NEON components will vary based on the needs of the facility. In this section we have tried to describe the most important issues to be considered for networking for NEON.

Networking within Observatories

NEON observatories will consist of several components, including a main site and satellite sites, each of which may contain a variety of data acquisition and data management equipment. Networking capabilities will need to scale to the requirements of each component, although it is clear that the main site will need much higher network capabilities in order to accommodate information from its affiliated satellites.

Main site networking will be generally supported by existing institutional infrastructure at observatories. The local area network (LAN) at the main site should include high-throughput networking, supporting at least switched 10Mbps Ethernet, and probably gigabit Ethernet. This will allow scientists working at an observatory to easily transfer data and images across the LAN for various scientific purposes. Data acquisition components and sensors can be directly wired to this site LAN or can be connected through a wireless network at the observatory (see below).

Wireless networking

All of the sites collecting sensor data over a large area should incorporate wireless networking technologies to aid in data collection. The network topology must reflect redundancy whenever possible. The backbone should be built using radios that use a widely accepted protocol (e.g., currently 802.11b) to insure interoperability. Each backbone node, in addition to being a point-to-point repeater, will also act as an Access Point for sensor arrays and scientists in the field using computers and personal digital assistants. The wireless access points will provide an Ethernet bubble over the research landscape that will connect multiple sensors to the backbone and researchers to the Internet and/or local database and modeling resources, even when they are located in remote field locations.

The network needs of a scientist in the field with a handheld computer (bi-directional, real-time, interactive) are very different from lab instruments and high resolution field sensors (uni-directional, non-interactive) and re-configurable field sensors (bi-directional, interactive). . The network design may need to factor in these different use-cases in order to be efficient.

Mobile wireless repeater modules would be used to extend wireless coverage to areas that need only temporary connectivity. An example would be seasonal or catastrophic event, such as flooding or a wild fire. The ability to quickly deploy these repeaters to areas outside of the normal wireless coverage and get real-time sensor data would greatly extend the utility of the network.

The devised data collection system should be tiered to allow for maximum density of coverage within a site. The primary tier would be a backbone wireless network encompassing the entire observatory. At each of the backbone nodes there would be a smaller regional node that extends the network at a finer grain (e.g., to extend into a river valley). This node then could act as an access point for low-powered sensor arrays using emerging communication protocols and schemes.

Networking satellite sites

Satellite sites will vary among observatories in their role and capabilities, so networking decisions for satellites must be flexible. For those satellite sites that serve to extend the data collection area (e.g., field stations), a network topology that is similar to the wireless network for the main site is probably appropriate. For satellites that perform sample processing or curation, or data processing, a traditional wired LAN network is probably adequate.

Each satellite site should be connected to the main site at a bandwidth that is appropriate for the data that needs to be transferred. Leveraging Internet or Internet2 connections for this is an obvious route, although there may be cheaper solutions that provide higher bandwidth via wireless networking depending on the spatial relationship to the main site. The HPWREN network in southern California demonstrates how wireless can provide very high bandwidth connectivity over a wide region at low cost.

Existing network connectivity at satellite sites will vary greatly from dial-up modems across phone lines to broadband connectivity. NEON will probably need to support a minimum bandwidth in order to keep satellite sites on par with the rest of the NEON sites. The baseline speeds (400 kbps download, 150 kbps upload) that are currently available from commercial satellite providers (e.g., Starband, Direct PC, Tachyon) are probably reasonable, although much higher bandwidth will be warranted for many satellites.

Telemetry & sensor networks

The core data needed from NEON sites will necessitate a wide variety of sensors in the field that capture, for example, meteorology, hydrology, and biodiversity data and imagery (still and video). Networking sensors so that their data are available without involving a technician, and possibly in real-time, will save a tremendous amount of labor. Consequently, networking sensors should be a requirement for most data collection activities. Leveraging the wireless network described previously for this is ideal, although a variety of practical considerations (e.g., power) will need to be considered at each observatory.

The core site should have the best connectivity and most diverse set of sensors and the bandwidth needed to distribute data in real time. Archiving of real time values is seen as important to insure complete data sets.

Sensor data collected in the field will be stored on the data loggers in the field and may be simultaneously transmitted and stored on site in a local data repository and at a core hosting site, probably at the main observatory site. Automated QA & QC should be performed; it may be optimal to do so at the satellite level before the data are sent to the main site for storage, although the overall NEON versioning and archiving protocols will need to be consulted.

Some sensors will produce large data streams, and so a mechanism for prioritizing and scheduling data flows from field stations to the satellite and/or main site may be needed. Nighttime delivery of large archival data sets will help balance out load on local networks.

Accessibility, travel time**,** spatial extent of sensor coverage, and heterogeneity of instrumentation will determine the number of personnel necessary to maintain sensor stations. Workshop participants were concerned that the core support requirements for data collection and sensor maintenance (as described by previous NEON workshops) far exceed personnel allocations that are possible under budget descriptions. The scientists involved in determining core data collection standards for NEON will need to be careful to consider staffing needs when choosing the core data to be collected at observatories.

We hope that NEON observatories can leverage existing personnel at observatories and distribute NEON personnel as partial

FTEs across satellites. Observatories should be able to create shared positions and overlapping job tasks in order to bridge gaps, but a stringent focus on data quality should still be maintained. We noted that some types of sensors (e.g., eddy flux towers) require dedicated, more highly skilled support than others in order to insure quality data.

Networking among Observatories

NEON will need to have a tightly coupled, reliable, and high-bandwidth network for transferring data and distributing computation across sites in the network and to the broader research community. It is critical at a national scale that the network be robust to failures at various points, including the loss of connectivity to one or more observatories. This necessitates a well-considered design for replicating data and services across nodes in the network for failover, load balancing, computational efficiency, and search efficiency. The Coordinating Unit would be primarily responsible for the design, implementation, and maintenance of the NEON wide area physical network.

As at the observatory level, leveraging Internet or Internet2 connections for the physical network is an obvious route, although there may be cheaper solutions that provide higher bandwidth via wireless networking depending on the spatial relationships of the sites. It would be a enabling accomplishment for NEON to begin building upon the HPWREN network in southern California to provide national wide-area network coverage for the NEON enterprise.

Each observatory will require a minimum dedicated bandwidth of 1.544 Mbps to the Coordinating Unit, but higher bandwidth (Intenet2) connectivity will undoubtedly be needed for many types of data transfer (e.g., images). The network topology will need to be determined through a careful analysis of how data will flow and where services will be provided, considering reliability and scalability. A hub-and-spoke model is probably not sufficient because of the inherent likelihood of point failures, so various ring topologies should be considered.

The NEON Coordinating Unit will be responsible for researching and recommending observatory networking solutions. They will take an active role in generalizing and scaling up local site solutions so they are applicable to other observatories in NEON, and to propagate positive experiences to the entire network. They should capitalize on the inherent strengths of individual observatories when developing new sensors and power and communications grids. Standard configurations for networking equipment should be used across the entire network where it is sensible and improves deployment efficiency (e.g., bulk purchases). The Coordinating Unit should also investigate cooperation with other local agencies (i.e., agreement with USFS, etc.) for placement of sensors and network relays on public lands.

Serious consideration should be given to whether or not to contract out for external management of the NEON wide area network. Many companies (e.g., CGNet) can build infrastructure for non-profit organizations and monitor and report data traffic patterns, indicating bottlenecks and surplus bandwidth. These types of features could be more cost-effective than managing a large suite of sites individually.

Networking to the Broader Research Community

One of NEON's principal goals is to fundamentally improve the nature, extent, and quality of data available to the broader environmental research and policy communities. Consequently, network provisioning that is targeted outside of the network needs to be carefully considered.

Researchers outside of the network should be able to discover and access data from a central access point that scales to support high bandwidth needs of researchers and policy makers. No matter how large the NEON network is internally, its success will be measured by its impact on the external users of its data. Again, we envision the Coordinating Unit as having primary responsibility for development of these external interfaces, and that they would largely utilize Internet2 for deployment of the network. However, in addition to web-based access, NEON should carefully consider the need for the enhanced capabilities provided via Grid-based access to NEON resources.

IX. Remote Sensing and Geographic Information Systems

Remotely sensed imagery must be an integral component of the core monitoring program for a NEON observatory. The network should possess a high level of expertise and be able to contribute significantly to advances in the management and interpretation of image data. In addition, the accurate geo-referencing of all field data collected through a NEON facility is essential to the ability to understand the spatial context of ecosystem function and the scalability of these observations to broad, regional questions. To ensure that both of these standards are met, the workshop participants identified a number of points to consider in developing an overall program design.

Data acquisition

Most of the image data used in NEON is expected to come from airborne and satellite sensors operated by independent agencies. Bulk purchases and comprehensive use agreements will need to be negotiated on a network wide scale to reduce costs and ensure consistent data coverage across the entire monitored area. The NEON network should work as a whole to negotiate cooperative agreements for data access that take into account the added value that can be contributed by the NEON network through long term ground truthing. The workshop endorses the suggestion made in a previous workshop that NEON seek a participation in EOS.

NEON observatories should establish an explicit set of guidelines with respect to the geo-referencing of field locations for all data observations, curated samples, and voucher specimens. These should specify metadata requirements regarding the measurement and recording of positional accuracy (at the feature level if necessary), explicit documentation of the coordinate system for the geometric data, and spatially explicit description of sampling parameters including sampling universe, stratification, and collection units. The workshop was not convinced of an imminent need for acquiring a dedicated sensor platform as opposed to negotiating with existing resources for NEON-targeted missions. However, the argument raised by a previous workshop that certain critical sensors are simply not available was acknowledged.

Processing

Resources must be provided for processing geospatial data in a consistent and timely (possibly real-time) manner. Some standard processing functions such as haze removal, rectification, standard indexes such as NDVI, etc. might be more efficiently done at a central facility using data from local and regional sources. Other processing, such as land cover classification, is expected to be dependent on algorithms developed on a regional basis and will require resources and expertise at the individual observatory facility. Providing ground truthing data for both local and regional classification needs is expected to be an integral part of the core monitoring program and will require standard policies for incorporating its collection in routine field activities.

Storage and management

The storage needs for long-term remote sensing data for the overall NEON observatories will be very high, far exceeding what is currently experienced within current networks (e.g., LTER sites). Even if long-term archive needs are met by either a central NEON office or through agreements with non-NEON partners, there are still significant needs at the observatory for storage of image data that far exceed what is currently available at most LTER sites today. Metadata formats must seamlessly accommodate both spatial and observational data, and storage solutions must be designed to ensure the long-term persistence of the links between observations and spatial geometries.

Research integration

NEON observatories should provide advanced support for the integration of spatial data within the overall research program. This requires several resources: (1) software interfaces to aid access to both raw data and to a comprehensive set of standard derived products such as classifications and indexes; (2) a distribution system that uses a network-standard programming interface to allow this software to access data from remote services that provide basic processing functions like subsetting, resampling, etc. prior to retrieval; and, (3) extensive GIS support staff to provide training, research support, and data analysis/processing.

Priorities for Archiving

As with other data types such as multimedia and simulation model output, remote sensing and grid-based analysis are expected to produce high volumes of data and the NEON network should develop guidelines for setting priorities on data retention. Two issues should be addressed: (1) standards for smart sensors that can sample or aggregate data as it is collected and (2) criteria to guide researchers in determining the relative importance of data products derived from multiple processing steps. The cost benefits of rolling lower-priority data to offline storage media have been shown to be offset by decreasing costs of additional online storage, requiring some thought be taken to how to set priorities.

X. Computation, Analysis, and Synthesis

The provision of advanced and standardized approaches to computation, analysis, and synthesis is one of the key advantages of NEON, and will have a major beneficial impact for NEON-affiliated researchers, as well as the general ecological community.

Many ecological and environmental scientists do not currently participate in a broadly shared culture of computing, modeling, and analyses. Their information technology usage is characterized by strong individualistic traditions, with a resulting variety of non-optimal computational solutions in terms of both hardware and software choices, as well as a diversity of methodological and analytical approaches. Much of this diversity is attributable to highly parochial traditions of usage and support, often with no direct consideration of objective computational criteria such as power, efficiency, or interoperability. We believe affordability and familiarity are often the primary criteria driving investigator's computational solutions.

NEON provides an outstanding opportunity to identify and overcome weaknesses in the reigning culture of computation, by providing advanced services, support, and leadership for a next-generation approach to analysis, modeling and synthesis for the environmental sciences. This must be accomplished in a way that enables current researchers to *easily* transition into a vastly more powerful and interoperable computing community.

General Computational Needs

The generation of new scientific insights from the NEON framework will heavily depend on the computational capabilities and analytical tools that are available to researchers. NEON should enable a new paradigm for accomplishing analyses within the environmental sciences, by addressing and solving several major issues that currently retard scientific progress. These issues include:

- Ability to effectively discover and analyze massive, distributed data sets
- Ability to access sufficient computational cycles for accomplishing highly complex or demanding tasks
- Access to scientific and analytical computational support and training, to assist and educate the user community in the new techniques and services enabled by NEON

In order to efficiently address these issues, we recommend that there be a well-defined, well-integrated set of NEON services relevant to analysis and modeling, including:

- Adoption of emerging GRID-based approaches as a unifying computing environment for NEON—including the development of well-defined protocols for intercommunication among NEON servers and services
- Standardization and parallelization of broad impact models and analytical services
- Provision of high-performance compute servers on which to run the above services
- Coordinated development of advanced visualization techniques to facilitate data mining and knowledge discovery within the NEON framework
- Development and support of advanced videoconferencing and other collaboration techniques for ecological science
- Development of standard hardware and software models for "NEON workstations"--that promote the use of “certified” applications and services throughout NEON (e.g., gcc, R, MPI, SSL, openGL)
- Promotion of developing and sharing models, functions, and scientific libraries based on select software systems that are scalable, and NEON-compliant
- Development of standard interfaces for visiting or non-aligned systems to view/access NEON services

We envision the NEON Observatories and Coordinating Unit as a well-connected, highly-integrated set of services and computational capabilities. GRID services provide an effective way to integrate NEON sites, as well as tap into a growing, generalized framework for scientific computing. This will be important in enabling NEON researchers to better access scientific data from outside NEON per se, as well as better expose NEON holdings to the general scientific, and particularly, earth sciences communities.

Adoption of GRID technologies should also enable NEON analyses to be run on vastly powerful, distributed computing servers. The ecological community in general has not taken full advantage of high-performance computing (HPC), and NEON provides the opportunity to remedy this situation. There is an initial problem, however, that many ecological and environmentally oriented software programs, including models, analyses, etc., are not currently written in a format optimal for high-performance computing (HPC) environments. We see NEON as providing critical services in identifying and enabling select models and analyses of significant community interest, so that these can take full advantage of GRID computation. It will also be necessary for NEON researchers to have transparent access to sufficient HPC resources to run these newly enabled services.

There are areas in which increasing computational capability and network bandwidth are creating new methods for accomplishing ecological, and in particular, synthetic science. These include advances in scientific visualization, an area that is traditionally under-utilized in ecological investigations. Given the vastly increased data sets and analytical capabilities within NEON, it will be necessary to create new methods for discovering and accessing these. We believe scientific visualization will be an important component of this, and feel that NEON should specifically facilitate domain-relevant developments in this area.

Similarly, synthetic science can be greatly assisted by emerging collaboration frameworks. These are likely to be standard services carried over the Internet, but here NEON can provide services (videoconferencing hubs, data-streaming, etc.) and support to create an ongoing virtual community of collaborating scientists.

Visiting researchers at NEON sites will frequently be most comfortable working with their own familiar and preferred hardware and software solutions. Thus, a well-defined interface must be developed for visiting and external researchers to achieve effective access to NEON services. The goal of this interface should not only be to provide access to NEON services by scientists, but also to enable external information systems to easily achieve compliance by implementing an application programming interface and thus participate more fully in the NEON framework. This extensibility will encourage broader community participation, with the possibility of evolving to a more GRID-based model for universal access to all computing services provided under NEON.

Distribution of responsibilities

Analytical and modeling capabilities provided by the NEON system should be superb and readily accessible to NEON researchers as well as publicly through the Web. NEON should provide transparent access to tiered analytical services, ranging from the desktop, to the Observatory server level, to high-end capabilities at the Coordinating Unit, and finally through strong liaisons with HPC Centers for the most demanding integration, simulation, and visualization needs.

The Coordinating Unit should take a strong role, in close conjunction and consultation with the individual observatories, in architecting and facilitating these NEON-wide services.

In order to promote availability and fault-tolerance, we recommend that each Observatory have replicated services with at least one other Observatory, or that these adopt some more robust solution promoted and supported by the NEON Coordinating Unit.

The extent to which the NEON Coordinating Unit would serve as a computational resource was widely debated within this working group. Although there was general acknowledgement that certain NEON-relevant computing services might be most efficiently hosted at a site specifically charged with those responsibilities, there was concern about which services and to what extent a single site (such as the Coordinating Unit) should be charged and equipped to handle these.

The following recommendations represent consensus opinion among the four subgroup members charged with addressing these issues, but in the broader forum a few individuals expressed the need for further consideration of how and whether to centralize any NEON computational services.

The primary role of the Coordinating Unit relative to NEON computation is seen as coordinating, developing, promoting, and deploying services that are consistent throughout the system, and especially where site-based replication of effort is deemed wasteful or redundant.

Another primary role would be to present NEON as a highly available data store and portal to computing services, with consistent and polished presentation of NEON-wide analytical services for both the public and scientific communities.

The Coordinating Unit could be assigned a number of additional functions that might best be undertaken at a single point of responsibility. We only cover activities relevant to the focus of this section—namely, analysis, modeling, and computation. These include:

- Design, implement and support a single security infrastructure for authentication and access
- Technical liaison between observatories and HPCC
  - GRID portal compliance of NEON services
  - Consultation on HPC issues and programming
- Facilitate collaborative analyses—shared accounts, data spaces, email lists, Web spaces, etc. (related to above)
- Central gateway for collaboration technology (dedicated teleconferencing room)
- Advanced virtual reality center for public outreach/display, education, and scientific insights (remote sensing applications, land-use)
- Implement and support services of NEON-wide relevance:
  - Coordinated calendar/schedule server with individual observatories
  - Core software and hardware licensing and negotiations
- Technology testbed and primary development site for new NEON-wide services
- Coordination of hardware and software contracts, acquisitions, and licenses
- Server aggregation site for accomplishing synthetic work that requires more significant compute servers ($100K-200K) than available at individual Observatories

Consultation and Training

We believe that the success of NEON's innovative computational approaches will heavily depend on scientists having easy access to an expert staff of qualified consultants who can directly assist and advise them in the adoption and use of NEON services. Moreover, we believe that a series of on-going training sessions will be necessary to achieve broad outreach and adoption of this radically new way of accessing and analyzing ecological data. Some examples that will require consultation include:

- Need for personalized consultation to develop customized data entry forms so individual data can be compatible with NEON information systems
- Generalized assistance-- On-line access to experts in the specific types of analyses and services provided by the NEON
- Specific assistance-- As needed access to data integration experts and knowledge engineers who can assist in accomplishing specific analytical tasks within NEON that might have general utility (e.g., integration of disparate forms of data; location of relevant observations from among vast data stores, etc.)

Without such assistance and training, we believe that much of NEON will be underutilized due to its novel and extensive capabilities.

Specific Computational Needs

We differentiate what we consider baseline computational needs and services provided at Observatories, from those at the Coordinating Unit. Our model is one of ample access to capability regardless of site location, with the emphasis on support for site-specific, specialized needs addressed primarily at the individual Observatories, while NEON-wide IT needs and analyses are the primary responsibility of the Coordinating Unit.

Observatory

Data preparation: Each individual Observatory must be able to solve its own challenges relative to local data ingestion, integration, aggregation, QA/QC; and the summarization and processing of these raw data streams for conformance to NEON standards, as well as preparation for downstream analysis. This will minimally require, at each Observatory site, a dedicated data concentrator server for receiving and housing real-time streams (e.g., from sensors), and a data preparation server for human assisted and pre-existing databases and datasets.

Desktops: for "conventional" analysis and modeling-- Desktop use of commercial off-the-shelf software (COTS) would be a starting point (e.g., MATLAB, SAS, Mathematica, ArcGIS, etc.).

Computer server: as a baseline a shared, multi-user compute server at each Observatory (~$20K-100K) running COTS; also possibly specialized, higher-end software. It is likely that sites will have specialized needs for esoteric or enhanced solutions in certain areas (e.g., for analyzing and visualizing hyperspectral imagery, or microarray processing). This may require purchasing specialized hardware and software, which should be decided upon in view of already existing services elsewhere within the NEON system.

Collaboration technologies such as videoconferencing and application sharing should be supported and encouraged at all Observatories. We recommend a dedicated teleconferencing center be located at each Observatory. We believe the Coordinating Unit can serve as the main hub for advising and directing these services throughout the NEON.

Note that some provision should be made at each site for administrative computing support. Aside from desktop configuration, and maintenance, and application assistance for administrative support staff, this might also include developing and maintaining an organizational database and individualized Web presence. The extent to which these administrative support functions can be centralized should be examined.

Documentation

NEON should develop, adopt, or support efforts to capture all NEON-supported analyses in well-documented forms that can be used for automated processing. Recent developments in standardized domain-specific metadata frameworks, and abstraction of analysis and modeling steps as “components” of a pipeline or labeled graph, should be closely followed for providing a robust and replicable documentation of analytical methodologies.

Automated metadata capture of specific data resources and analytical procedures deployed, should be a top priority in development of all NEON-based tools wherever possible, and these should comply with or constitute standards for the larger ecological community.

Software

NEON should prioritize development of software for two specific purposes: to enable domain access to advanced GRID services such as data discovery and visualization, and analytical operations upon high-performance hardware; and the specification and building of a model NEON workstation based on select standardized software components. We recommend the use of COTS wherever possible, although we endorse a direct focus of training and support on a carefully chosen subset of currently available options.

XI. Personnel

Workshop participants attempted to clarify the staffing needs for information management at NEON at both the observatory and Coordinating Unit. The workshop identified a number of specific needs at both the network and observatory level. However, given budget uncertainties and the uncertainty about the scope of data collection, we were unable to create a precise roster of staff needed. Instead, we were able to construct a list of the roles to be filled at a NEON observatory and at the Coordinating Unit, noting that how these are filled will partly depend on the role that is ultimately specified for the Coordinating Unit.

Observatory Personnel

Observatory staff (see Table 1) should participate actively in network-wide standing committees to ensure complete and uniform participation in the development of standards and procedures for data management across the network.

Staff should play a significant role in providing training for researchers using NEON facilities. This may involve training orientations, developing online documentation, and protocol guides. There should also be funds for graduate student positions in NEON. Students can be trained in informatics techniques while simultaneously providing informatics research support to research scientists.

Participants felt that it was critical for each observatory to have a dedicated information manager who managed all aspects of the data systems. This person would supervise a variety of other support personnel, including system administrators, desktop computing support staff, systems developers, GIS and analytical support staff, GIS specialists, and field service technicians.

The Systems Integrator role was considered the primary means to incorporate new technology at the observatory. This person would interact closely with development engineers at the Coordinating Unit, integrating and deploying systems in a production environment.

Field service technicians are critical to collecting data and maintaining field networks, sensors, and equipment. The exact number of and skills for these technicians depends on the scope of data collection and a detailed assessment of the travel time to various remote sensor locations. We envisioned a need for individuals with expertise in wireless networking, flux towers, hydrological sensors, climate sensors, and other specialized equipment, as well as field crews for collecting survey and inventory data and monitoring experiments. This could represent a large resource allocation depending on the scope of core data collection within NEON.

Table 1: Observatory personnel

PI for informatics	1
Information Manager	1
System/Network Administrator	1
Desktop Support	1-2
Systems Integrator/Developer	1
GIS Specialist	1
Quantitative analyst	1
Field Service Technicians	5?

Coordinating Unit Personnel

The personnel for the NEON Coordinating Unit would depend heavily on the actual role and responsibility that the unit assumed in the network. Coordinating Unit level personnel should participate in the information management committees described above, be specifically charged with informing the network of industry developments, assume a leadership role in the implementation of standards and procedures agreed upon by the network, and assume a service role in the development of software tools that assist this process.

Table 2: Coordinating Unit personnel

Director	1
Assistant Director for Informatics	1
Development Engineers	5-8
Site Deployment Support	2
Web Administrator	1
Systems and Network Administrator /Desktop Support	1
Administrative Support Staff	2
Informatics researchers	

The roles at the Coordinating Unit focus on the development of information management systems for use across the entire network and on support of individual observatories in deploying this infrastructure. Consequently, we emphasized a need for a small group (5-8) of development engineers that would develop the core systems and interfaces for the network. These individuals would focus on developing systems that meet practical needs of the network and observatory, deploying these systems, and stability of these production systems. They would not be involved in cutting-edge informatics research.

Support personnel, such as the web and system administrators, would maintain network-wide computing services such as databases, web portals, directories, and computational clusters.

The Site Deployment staff would include people that could travel to sites to assist in deploying systems at observatories and satellites, and would help to resolve technical problems encountered when adapting systems to existing infrastructures at observatories.

The informatics researchers are a group of people pursuing advances in computer science and information technology (e.g., in web services, semantic mediation, knowledge engineering). They would probably be externally funded through agency research programs. We included them here to point out that it is critical that researchers in NEON push the informatics research agenda, making sure that new advances in computer science are relevant to NEON and expand information technology capabilities in NEON. Co-locating them with NEON development engineers would have a synergistic effect, allowing NEON to take advantage of new research advances quickly and efficiently.

XII. Summary

Participants at this sixth NEON workshop were excited about the scientific advances that NEON would enable. We were generally supportive of the concept of building a tightly integrated research platform that can be used for the wide variety of ecological studies needed for our society. This report outlines the issues that would need to be considered in implementing information management systems for NEON. Foremost among those considerations is the need for advanced planning of data collection standards and network-wide information management procedures under realistic budget scenarios. Consequently, for NEON to realize its true potential, it is imperative that the advance planning to make NEON a **network** occurs before funding individual observatories, and that all funded observatories fully endorse the concept that they are but one node in a larger entity. One of the best ways to accomplish this is to ensure that the NEON Coordinating Unit plays a strong role in the network from the very beginning.

XIII. Appendix: Workshop Participants

Sandy J. Andelman
National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara
735 State Street, Suite 300
Santa Barbara, CA 93101-5504
USA
TEL (805) 892-2505
FAX (805) 892-2510
email andelman@nceas.ucsb.edu

James H. Beach
Natural History Museum and Biodiversity Research Center
University of Kansas
1345 Jay Hawk Boulevard
Lawrence, KS 66045
USA
TEL (785) 864-4645
FAX (785) 864-5335
email jbeach@ukans.edu

Barbara Benson
Center for Limnology
University of Wisconsin, Madison
680 North Park Street
Madison, WI 53706-1492
USA
TEL (608) 262-2573
email bbenson@lternet.edu

Pablo J. Bryant
Field Station Programs, College of Sciences
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-4614
USA
TEL (619) 594-8025
email pbryant@sunstroke.sdsu.edu

Scott L. Collins *
Division of Environmental Biology
National Science Foundation
4201 Wilson Blvd., Room 635
Arlington, VA 22230
USA
TEL (703) 306-1479
FAX (703) 306-0367
email scollins@nsf.gov

Deborah Estrin
Center for Embedded Networked Sensing
University of California, Los Angeles
3440 Boelter Hall
Los Angeles, CA 90095
email destrin@cs.ucla.edu

Jim Frew
Donald Bren School of Environmental Science and Management
University of California, Santa Barbara
6715 Ellison Hall
Santa Barbara, CA 93106-5131
USA
TEL (805) 893-7356
FAX (805) 893-7612
email frew@bren.ucsb.edu

Jeffrey Goldman *
Infrastructure for Biology at Regional to Continental Scales
American Institute of Biological Sciences
1444 Eye Street, NW
Suite 200
Washington, DC 20005
USA
TEL (202) 628-1500 ext. 225
FAX (202) 628-1509
email jgoldman@aibs.org

Don Henshaw
Pacific Northwest Research Station
USDA Forest Service
3200 Southwest Jefferson Way
Corvallis, OR 97331
USA
TEL (541) 750-7335
email dhenshaw@lternet.edu

Matthew B. Jones
National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara
9087 Sheiye Way
Juneau, AK 99801
USA
TEL (907) 789-0496
FAX (425) 920-2439
email jones@nceas.ucsb.edu

* indicates workshop observer

John Marchioni *
Gordon and Betty Moore Foundation
USA
TEL (415) 561-7566
email John.Marchioni@moore.org

Peter McCartney
Center for Environmental Studies
Box 873011, Arizona State University
Tempe, Az. 85287-3011
USA
TEL (602) 965-6791
email peter.mccartney@asu.edu

William K. Michener
LTER Network Office
University of New Mexico
801 University SE, Suite 104
Albuquerque, NM 87106
USA
TEL (505) 272-7831
email wmichene@lternet.edu

William Piel
Department of Biological Science
University at Buffalo
608 Cooke
Buffalo, NY 14260
USA
TEL (716) 645-3153
email wpiel@buffalo.edu

Tom Prudhomme
National Computational Science Alliance
National Center for Supercomputing Applications
605 East Springfield Avenue
Champaign, IL 61820-5518
USA
TEL (217) 244-3425
email tip@ncsa.uiuc.edu

Arcot Rajasekar
San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
USA
TEL (858) 534-8378
FAX (858) 534-5077
email sekar@sdsc.edu

* indicates workshop observers

O. J. Reichman
National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara
735 State Street, Suite 300
Santa Barbara, CA 93101-3351
USA
TEL (805) 892-2504
FAX (805) 892-2510
email reichman@nceas.ucsb.edu

Mark P. Schildhauer
National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara
735 State Street, Suite 300
Santa Barbara, CA 93101-3351
USA
TEL (805) 892-2509
FAX (805)892-2510
email schild@nceas.ucsb.edu

Wade Sheldon
Department of Marine Sciences
University of Georgia
Athens, GA 30602-3636
USA
TEL (706) 542-5955
FAX (706) 542-5888
email sheldon@arches.uga.edu

Sylvia Spengler
National Science Foundation
USA
email sspengle@nsf.gov

Greg Stossmeister
Joint Office for Science Support
University Corporation for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307-3000
USA
TEL (303) 497-8692
email gstoss@ucar.edu

John Weiczorek
Museum of Vertebrate Zoology
University of California, Berkeley
3101 Valley Life Sciences Building
Berkeley, CA 94720
USA
TEL (510) 643-0352
FAX (510) 643-8238
email tuco@socrates.berkeley.edu