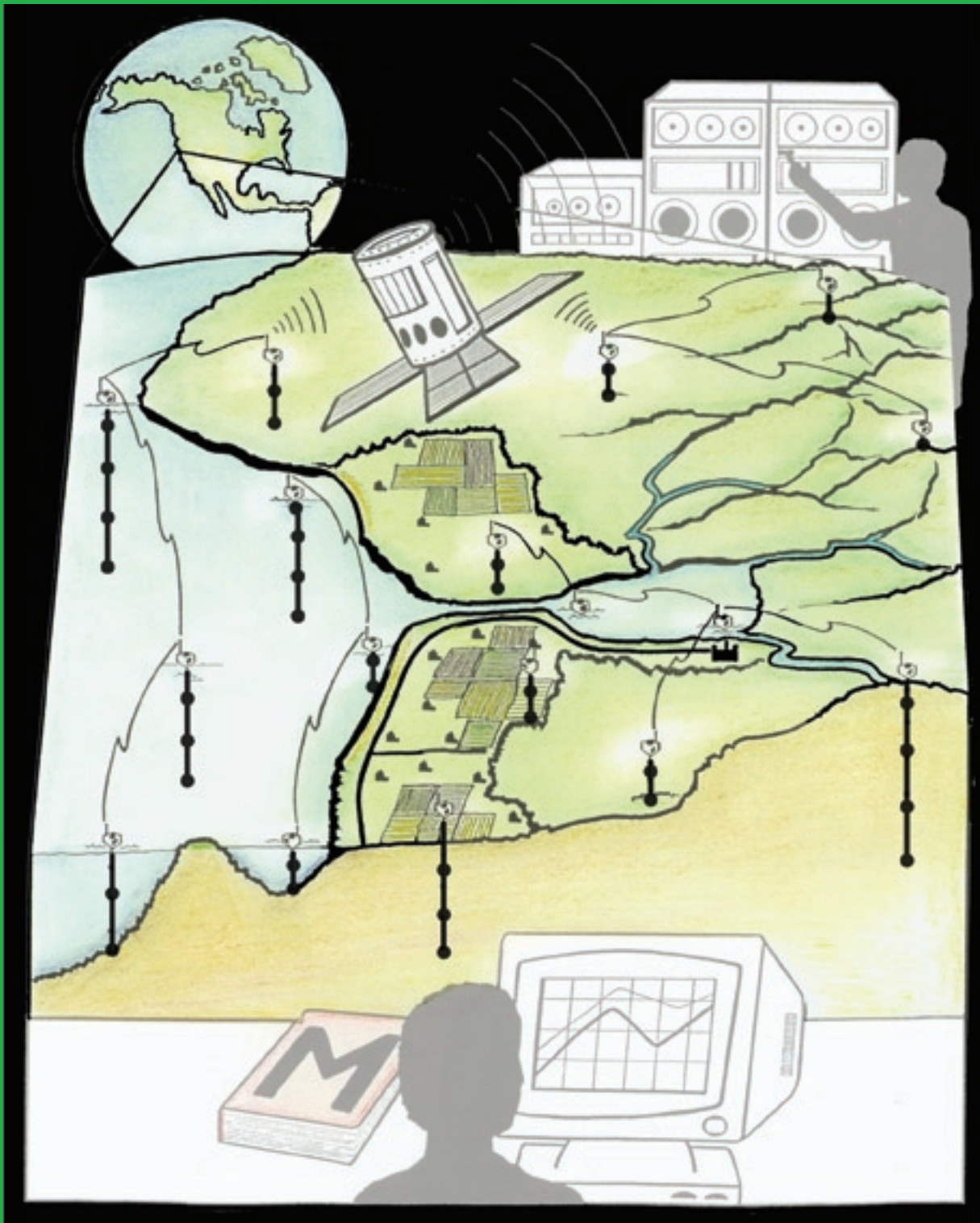


Environmental Cyberinfrastructure Needs For Distributed Sensor Networks

A Report From a National Science Foundation
Sponsored Workshop

12-14 August 2003, Scripps Institute of Oceanography



Environmental Cyberinfrastructure Needs for Distributed Sensor Networks

A Report from a National Science Foundation Sponsored Workshop

12-14 August 2003

Scripps Institute of Oceanography

Deborah Estrin

Center for Embedded Networked Sensing (CENS)
University of California, Los Angeles

William Michener & Gregory Bonito
Long Term Ecological Research (LTER) Network Office
University of New Mexico, Albuquerque

And the Workshop Participants

This workshop was supported by the National Science Foundation under grant number 0332827. All opinions, findings, conclusions, and recommendations in any material resulting from this workshop are those of the workshop participants and do not necessarily reflect the views of the National Science Foundation.

Estrin, D., W. Michener, G. Bonito, and the workshop participants. 2003. *Environmental Cyberinfrastructure Needs for Distributed Sensor Networks: A Report from a National Science Foundation Sponsored Workshop*. Scripps Institution of Oceanography, La Jolla, CA. 12-14 August 2003.

Table of Contents

Executive Summary	5
--------------------------------	---

Chapters

Chapter 1 —Introduction: Workshop Rationale and Scope	8
Chapter 2 —Sensing Technology	12
Chapter 3 —Deployed Sensor Arrays	20
Chapter 4 —Error Resiliency	26
Chapter 5 —Security	34
Chapter 6 —Data Management	40
Chapter 7 —Metadata	46
Chapter 8 —Analysis and Visualization	54

Boxes

1. A Vision for Environmental Sensor Networks	8
2. CUAHSI—Consortium of Universities for the Advancement of Hydrologic Science, Inc.	11
3. Research Experience in Self-organizing Networks: Motes, TinyOS, and TinyDB	15
4. GEON—the Geosciences Network	15
5. SpecNet—Spectral Network	17, 18
6. Embedded Networked Sensing	25
7. NSF CLEANER Initiative—Collaborative Large-scale Engineering Analysis Network for Environmental Research	29, 30
8. Fixed Ocean Observatories	37, 38
9. NEON—National Ecological Observatory Network	41, 42
10. North Temperate Lakes Monitoring	53
11. Observing the Acoustic Landscape	58

Figures

4.1. Example of a reconfigurable protocol sensor platform	31
7.1. Metadata management as a core component of a sensor network	48
7.2. Metadata management in the scientific workflow	49
7.3. Registries for sensor networks	49

Tables

2.1. Examples of existing sensors	12
7.1. Achieving metadata ‘buy-in’	51

Appendix A. Workshop Participants	60
--	----

Executive Summary

Increasingly, spatially extended networks of multi-variable intelligent sensor arrays are seen as revolutionary tools for studying the environment. The temporally and spatially dense monitoring afforded by this technology portends a major paradigm shift in environmental science and engineering—enabling scientists to reveal previously unobservable phenomena. Realizing this vision will require new cyberinfrastructure capabilities, methodologies, middleware, deployed infrastructure and a community of multidisciplinary scientists and engineers equipped to pose newly-enabled scientific questions.

To better define the cyberinfrastructure challenges and to search for creative solutions, a workshop entitled “Environmental Cyberinfrastructure Needs for Distributed Sensor Networks” was convened at the Scripps Institution of Oceanography. Approximately 75 participants from the environmental sciences, engineering, computer science, statistics, and mathematics focused on the cyberinfrastructure challenges being experienced by existing and emerging environmental networks (e.g., Collaborative Large-scale Engineering Analysis Network for Environmental Research (CLEANER), Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), National Ecological Observatory Network (NEON), and Long Term Ecological Research (LTER) Network) as they implement distributed sensor networks. Such networks will play a crucial role in developing the databases and knowledge that support understanding of natural and human-dominated ecosystems as well as appropriate strategies for early warning and environmental remediation.

Workshop participants made several recommendations in relation to basic research needs, education, outreach, and collaboration and partnering. (Note: [#] refers to chapter(s) where recommendations are discussed in more detail.)

Sensors

- **Design more capable sensors.** Research should focus on sensor designs that support long-term integrity, performance, interactivity, minimal environmental impact and minimal power consumption. [2]

Deployed Sensor Networks

- **Invest in prototyping and end-to-end testbeds.** Sensor networks—including sensors, network security, information technologies—must be tested and validated in large-scale natural environments across a range of applications and target domains. Validation will require reasonably controlled experiments; i.e., comparing different sensor networks in the same systems, and running the networks side-by-side with traditional monitoring techniques to provide a form of ground-truth. [2,3,4,5,6,8]
- **Support tool development for:** automated system layout and coverage estimation; composition and configuration of synthetic and simple sensors, and validation and calibration of sensor systems. [4]

Cyberinfrastructure for Sensor Networks

- **Support a new genre of cyberinfrastructure research and development for scalable sensor arrays.** Cyberinfrastructure research related to middleware and services (e.g., time synchronization, localization, *in situ* calibration, adaptive duty cycling, programmable tasking, triggered imaging) for scalable sensor arrays is essential to achieve the leaps in capability needed for hyper-scalability, sustainability, and heterogeneity. [3,6]
- **Build the requisite Grid and Web services.** The integration of GRID-based systems and Web services to convert raw environmental data into information and, finally, knowledge will be critical as sensor networks become increasingly ubiquitous. [8]

Metadata

- **Support development of metadata tools.** Enabling tools must be developed and provided to the community—including wizards to assist in sensor description and tools for automated metadata and data encoding. [4,7,8]
- **Engage the community in standardization efforts.** Community stakeholders (sensor developers, users, informatics specialists, standards organizations) must be engaged in the metadata standardization process including the design, development, implementation, testing, and adoption stages. [7]

Security and Error Resiliency

- **Support cyber security research and development.** Cyber security solutions must be sought through research that will not hinder free and open exchange of most data, but that will also protect the network and its sensors as well as provide the ability to restrict access to highly sensitive data when needed. Research in enhanced security “middleware services” is critical given the limited resources available on a sensor node. [5]
- **Enhance error resiliency.** Support research and development of sensor network autonomic approaches such as self-diagnosis and self-healing that relieve the user from the burden of attending large numbers of nodes individually, as well as system level techniques that ensure resilient operation of the network in the presence of a small percentage of compromised nodes. [5]

Analysis and Visualization

- **Support algorithm development.** The design and operation of complex sensor network applications require new algorithms from statistics, machine learning, and visualization. New analysis and visualization tools should enable processing and interpretation of high-bandwidth sensor streams and target new functionality in knowledge discovery and dissemination. [8]
- **Enhance visualization capability.** It is necessary to enhance visualization tools on handheld mobile devices (PDAs, cell phones, etc.), as well as to develop new display systems that integrate high-resolution imagery and video, high-fidelity audio and tactile interfaces to support virtual and augmented reality environments. [8]

Education

- **Educate the next generation of computer scientists, engineers and domain scientists that will design, implement and deploy sensor networks.** Such training should be interdisciplinary to foster scientists and engineers that are well versed in the scientific and technical capabilities of sensors and sensor arrays and the databases resulting from their use, as well as the appropriate domain sciences and information technologies (including ethics and privacy implications, Web and Grid services for data and tools, and advanced analysis and visualization cyber

security). [2,3,4,5,6,8] Potential mechanisms include jointly-funded projects, development of interdisciplinary curricula and hands-on workshops. [3,8]

Outreach

- **Promote outreach to the public, decision-makers, and resource managers.** It is critical that we develop appropriate information systems and methods for providing the data and information collected by sensor networks in a form that is compelling and informative to policy makers and the general public. [3,4,5,6,8]

Collaboration and Partnering

- **Build partnerships.** Sponsored programs should strongly encourage multiple institution investigations (e.g., coupling universities with national research laboratories and industry), promote research by multi- and interdisciplinary scientific teams, and support strong collaborative linkages with standards organizations. [2,3,6,7,8]
- **Sustain long term deployments.** We need to develop collaborations and review processes that keep facilities alive, evolving, and non-obsolete. We also need funding models that recognize the importance of staffing for stewardship and management. [3]
- **Promote open source solutions and repositories.** Incentives for and ease of contributing to open source toolsets, models and testbeds are essential to shift the community toward developing reusable system components and enhancing interoperability. [3,4,6,8]

Chapter 1. Introduction: Workshop Rationale and Scope

The widespread proliferation of the Internet and other technological advances—particularly wireless and acoustic transmission from remote sensors, coupled with the decreasing cost, size, and weight of sensors—is resulting in a major paradigm shift in environmental science and engineering. **Spatially extended, intelligent networks of multi-variable intelligent sensor arrays are emerging as revolutionary tools for studying complex real-world systems. The temporally and spatially dense monitoring afforded by this technology promises to reveal previously unobservable phenomena.** An immediately attractive feature for researchers is the potential for remote manipulation of experiments or observing networks in near real-time based on incoming data from the local network, from nested or adjacent networks, or from remote sources.

Box 1. A Vision for Environmental Sensor Networks

Pervasive *in situ* sensing of the broad array of environmental and ecological phenomena across a wide range of spatial and temporal scales.

Sensor networks should be robust and autonomous, be inexpensive and long-lived, have minimal infrastructure requirements, and be flexible (expandable and programmable) and easily deployed and managed.

Sensor network data should be maximally self-documenting and of known quality, readily integrated with other sensor data, and easily assimilated.

The growing demand of the environmental science and engineering communities for such advanced sensor systems raises critical cyberinfrastructure issues, including:

- What are the optimal protocols for two-way communication with sensors and dynamic control of sensor networks?
- What are the most effective mechanisms and protocols for rapid data transmission?
- What is the best way to dynamically manage sampling schemes at nodes with limited power budgets when multiple sensors share the same power source?
- What is the best way to manage heterogeneous physical, chemical, and biological data streams that include both high-bandwidth streams (e.g., video data and broadband seismic data) and low-bandwidth streams

(e.g., temperature sensors)?

- What are the best methods to collect, manage, archive, and distribute data from such systems?
- What technologies provide the best access to remote computing resources for processing and visualization of the data collected?
- Where are the software tools for the analysis of the multidisciplinary, spatially extended, intermittent datasets that will emerge from such observing systems?
- Do we have appropriate knowledge representation software to ensure that these data are easily accessible and seamlessly shared across disciplines?
- What are the most reliable methods to insure the integrity of the communications and control systems for such observing networks, together with the integrity of the data management and archiving systems?
- How can researchers automate quality control of the data?

To realize the vision set forth in this report will require the development of cyberinfrastructure capabilities, methodologies, middleware, and deployed infrastructure, as well as a community of multidisciplinary scientists and engineers who are equipped to pose the new classes of scientific questions that will be supported by this technology.

Beyond the major advances that will be enabled within each discipline by this cyberinfrastructure, the new paradigm will open a broad range of opportunities—from scientific research and science-based policymaking to education—through the rich possibilities for wider access and cross-disciplinary uses of the data collected from these advanced sensing networks. For example, in environmental science the ability to compare data from multiple, apparently unrelated sensing networks will enable researchers to examine new phenomena and to decipher unanticipated interactions among systems in order to advance our understanding of the Earth's environment.

Other factors that arise in networked sensor systems include cyber security and intellectual property issues. Solutions must be sought that will balance the free and open exchange of most data, with protection of networks and their sensors, and the ability to restrict access to highly sensitive data when needed. Finally, the inte-

gration of end-to-end GRID-based systems that convert the raw environmental data into information and, finally, knowledge, will be an increasingly important activity as sensor networks become more ubiquitous.

To better define the issues outlined above and to search for creative solutions to these challenges, a workshop entitled “Environmental Cyberinfrastructure Needs for Distributed Sensor Networks” was convened at the Scripps Institution of Oceanography. The workshop was attended by approximately 75 participants (Appendix A) from a range of disciplines including the environmental sciences, engineering, computer science, statistics, and mathematics. A steering committee (Appendix A) comprised of members of the research community and spanning the participating disciplines designed the workshop format and organized efforts to invite participants.

The workshop included talks by William Michener from the University of New Mexico (“Environmental Sciences: Informatics and Infrastructure Challenges”), Deborah Estrin from UCLA (“Embedded Networked Sensing for Environmental Monitoring”), Michael Horton from Crossbow Technologies (“Commercialization of Sensor Networks”), Stuart Gage from Michigan State University (“Terrestrial Environmental Sensor Networks: Challenges and Opportunities”), and Frank Vernon from the Scripps Institution of Oceanography (“Real-Time Sensor Networks: Lessons from ROADNet and HPWREN”). These initial presentations provided background for workshop participants.

The remainder of the workshop was organized around seven directed discussions that focused on: (1) sensing technology; (2) deployed sensor arrays; (3) error resiliency; (4) security; (5) data management; (6) metadata; and (7) analysis and visualization. Deborah Estrin and William Michener proposed a vision for environmental sensor networks (Box 1) and the seven breakout discussion groups were charged with identifying the key activities necessary for the vision to be realized.

Participants were specifically charged with addressing four issues in each of the breakout sessions: (1) characterizing the nature of the problem to achieve the vision for environmental sensor networks; (2) identifying the most significant challenges that must be addressed; (3) proposing constructive solutions to overcoming those challenges and identifying the communities best poised to address a particular challenge; and (4) providing a concise list of recommendations for action.

For each of the seven directed discussions, one overarching question was tailored to initiate the discussion for each breakout session:

- **Sensing technology**—What are the greatest needs for sensor component development for the different communities represented?
- **Deployed sensor arrays**—What are the most urgent needs in relation to deploying sensor arrays in the field to achieve the overarching vision outlined above?
- **Error resiliency**—How do we best characterize and optimize data quality from systems composed of large numbers of noisy/faulty channels?
- **Security**—How can we construct flexible, light-weight systems that are not excessively vulnerable to denial of service or inappropriate access to data and resources?
- **Data management**—How do we best manage and archive high-bandwidth, heterogeneous physical, chemical, and biological data streams from fielded sensor arrays?
- **Metadata**—What metadata developments are needed to promote data discovery, access, integration, and synthesis?
- **Analysis and visualization**—What are the appropriate tools for analyzing and visualizing the complex, multidisciplinary, spatially extended data that will emerge from new and comprehensive sensing systems?

The results from the discussion groups comprise the remaining seven chapters of this workshop report. Each workshop participant was physically present for and contributed intellectually to two different discussion breakout topics. Three to five participants for each topic remained in San Diego for an extra day to write the report. Breakout leaders and reporters included:

Sensing technology:

Leader - Peter Mikhalevsky
Reporter - David Brady

Deployed sensor arrays:

Leader - Deborah Estrin
Reporter - Jeremy Elson

Error resiliency:

Leader - Tom Harmon
Reporter - Lewis Girod

Security:

Leader - John Stankovic
Reporter - Arthur Maccabe

Data management:

Leader - Stuart Gage
Reporter - Robert Stevenson

Metadata:

Leader - Frank Vernon
Reporter - Gregory Bonito

Analysis and visualization:

Leader - Tony Fountain
Reporter - Paul Flikkema

In addition to producing this report, the workshop has had a number of broader impacts. First, a multidisciplinary group of eight graduate students was able to participate in organizing the workshop and actively contributed to each of the breakout sessions, as well as the workshop report and related publications. The students had the opportunity to work closely with a diverse group of scientists from the many disciplines represented. Second,

six international scientists contributed significantly to the workshop, bringing their unique expertise to the discussions and participating in fruitful interactions with both graduate students and U.S. scientists during the workshop and during informal discussion periods. Third, the workshop served as an important venue for bringing together scientists and engineers from a diverse mix of disciplines (computer science, engineering, mathematics, biology, geology, etc., who would not normally collaborate) to address a common problem—the creation of cyberinfrastructure capable of supporting meaningful distributed environmental sensing to address critical science and engineering challenges. It is anticipated that a long-term benefit of the workshop will be to catalyze future collaborative research efforts among the many disciplines present.

The workshop also brought disciplinary expertise to bear on challenges that are being or will be faced by existing and emerging environmental networks such as the NSF Collaborative Large-scale Engineering Analysis Network for Environmental Research (CLEANER), the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), the National Ecological Observatory Network (NEON), and the Long Term Ecological Research (LTER) Network. These networks (see boxes throughout the report) will play an indispensable role in developing the databases and knowledge to provide vital understanding of natural and human-dominated ecosystems, as well as appropriate strategies for early warning and environmental remediation.

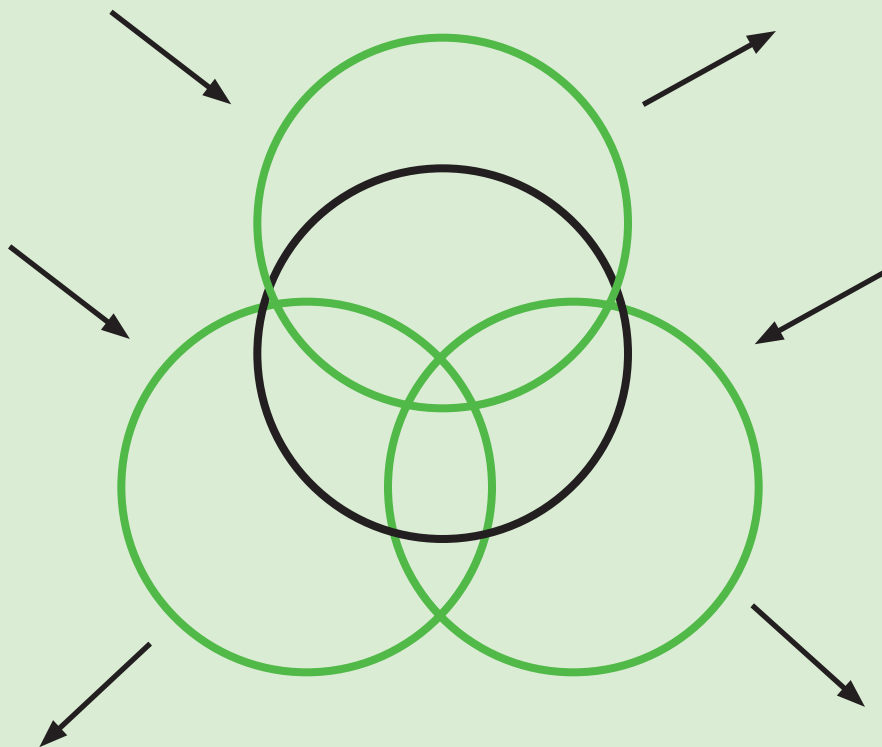
Box 2. CUAHSI—Consortium of Universities for the Advancement of Hydrologic Science, Inc.

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) (<http://www.cuahsi.org>) is an organization currently representing 73 US universities whose goal is to develop infrastructure and services at academic institutions supporting the advancement of hydrologic science. By working collaboratively through CUAHSI, the hydrologic science community can achieve a scale of investment in research infrastructure and accomplish goals that are beyond the reach of individual investigators or laboratories.

The CUAHSI vision contains four main research components, as shown in Figure 1: Hydrologic Observatories, Hydrologic Synthesis, Measurement Technology and Hydrologic Information Systems. Hydrologic Observatories are the central focus of this vision; each observatory will cover a region of approximately 10,000 km² and contain data collection and experimental systems to clarify critical hydrologic science questions. A key CUAHSI focus is on the development of a network of field observatories.

The “paper prototype” of an observatory (i.e., no field data collection) is being developed with a \$500,000 grant from NSF for the Neuse River in North Carolina and a separate grant has been obtained to develop a Hydrologic Information System (HIS) for the entire CUAHSI network. The HIS will be designed in coordination with other NSF efforts in the geosciences, such as the Geosciences Network (GEON), CHRONOS (<http://www.chronos.org/>) and Geochemical Earth Reference Model (GERM), as well as with the NSF National Science Digital Library (NSDL) and Digital Library for Earth System Education (DLESE) programs in digital library technology and earth science education respectively.

Figure 1. Principal CUAHSI hydrologic research programs





Chapter 2. Sensing Technology

2.1 Introduction

2.1.1 Sensor Types

A diverse array of sensor types and functions is needed if we are to meet the current and future needs of environmental networked sensing. This diversity poses considerable challenges to the communities involved in designing, deploying, and utilizing sensor networks.

Sensors can be classified in several ways. From the perspective of sensor design and operation, sensors fall into

two broad categories with respect to their operational and power requirements: active and passive. An additional way to categorize sensors is to define the different environments or phases that are to be sampled. For example, the requirements and design of a CO₂ sensor vary depending upon whether it samples in the liquid phase (e.g., marine or aquatic environment) or the gaseous phase (e.g., atmospheric environment). In general, there is a need to broaden the application and capabilities of sensors to work in such a diverse array of environments.

Table 2.1 Examples of existing sensors grouped by category (physical, chemical, and biological) along with examples and comments.

Sensor category	Examples	Comments
Physical	Temperature (e.g. thermistor, thermocouple, IR sensor)	Cheap, reliable, low power requirements, small size
	Moisture (e.g. dielectric constant sensors)	Intermediate in price
	Wind speed and direction (e.g. hot wire anemometers cup anemometers, sonic anemometers)	Expensive, less reliable
	Photosynthetic Active Radiation	Intermediate in price, low power requirements
Chemical	Nitrate sensor	Expensive, under development for long-term field use
	Carbon dioxide sensor	Expensive, moderate power requirements
	Phosphorus sensor	Not available for field <i>in situ</i> use
	Protein sensors	[under development]
Biological	Video camera	Can detect motion, but requires pattern recognition, high bandwidth, and moderate power requirements
	Microphones	Can detect and distinguish organisms, but requires pattern recognition, high bandwidth, and moderate power requirements
	Lidar, optical sensors	Can detect above-ground biomass, but requires calibration, and misses below-ground properties
	Species sensors, individual sensors (population counts)	Generally not available; must be determined by the investigator in the field, or inferred from other methods and sensors
	<i>In situ</i> DNA analysis	[under development for aquatic systems; Not available for terrestrial systems]

From the perspective of disciplinary science objectives, sensors can be grouped into three general categories: *physical*, *chemical*, and *biological* (Table 2.1). In general, physical sensors are the most developed and reliable under field conditions; chemical sensors tend to be less available and reliable; and many important biological sensors do not yet exist. Thus, researchers often sense physical properties or biogeochemical processes *associated with* forms of life, rather than directly sensing life itself. To date, key biological variables (such as species distributions or diversity, population densities, productivity, and migration speeds and patterns) generally require a continuing or frequent human presence. In “difficult” or inaccessible environments (e.g., marine or below-ground realms), unattended biological sensing is a particular challenge.

Alternatively, computing and computing network solutions can be applied to the problem (e.g., via intelligent sensors, networks, and analytical capabilities). To some extent, the lack of appropriate, direct biological sensors means that biological variables must be inferred from other (e.g., physical or chemical) sensors, and may require a high degree of intelligence applied to the sensing system. In this case, the network itself, and all of the analytical capabilities applied to the network (including the human observer), become the “sensor.” Thus, from a sensor or systems design perspective, the biggest challenges—and biggest need—may lie in the biological realm, with the chemical realm close behind. Given the lack of availability of direct biological sensors for field applications, which may continue for some time, there is a considerable challenge in defining how biological variables can best be inferred from a range of indirect physical and chemical sensors.

Often, some level of abstraction can assist in addressing these variables. For example, the concept of “functional type” can replace “species,” simplifying the task of detecting the presence or activity of individual organisms. Alternatively, the presence or activity of organisms can often be inferred from biogeochemical, thermal, or acoustic signals. Lessons from astrobiology (e.g., determining fundamental indicators of “life” and characterizing the range of potential biogeochemical indicators associated with various forms of life) could be particularly instructive in designing novel and useful biological sensors. In addition, advances in DNA methodologies, including molecular probes and gene arrays, offer much promise for direct biological detection and the potential of being incorporated into sensor arrays.

Networked sensing will provide a critical framework for testing novel approaches to the problem of biological sensing and detection. The ability to have large-scale

networked sensor arrays in place can stimulate development and improvement of physical and chemical sensors, and the development and application of novel biological sensors.

Sensors in a networked array, and particularly in autonomous wireless arrays deployed in unattended environmental settings, must meet additional and different requirements than sensors in a laboratory or stand-alone field settings. These different requirements include capabilities inherent to the sensor and capabilities that allow the sensor to function as a component of the array and to interact with other elements in an array.

2.1.2 Desired Features

Workshop participants have identified five areas in which networked sensors pose new requirements: integrity and reliability, power efficiency, interactivity, affordability, and environmental impact.

Integrity and reliability. Often, sensor networks will provide unique and essential means to monitor long-term aspects and activities of an environmental system. In such applications, all components of the array—but particularly the sensors—need to provide reliable, unattended performance over seasonal to inter-annual time periods. Sensor designs, including packaging options, need to maximize sensor durability and integrity, and to minimize material and performance degradation from environmental exposure, contamination, and fouling. For many environmental applications, throw-away sensors developed for commercial or military applications may not provide acceptable long-term performance characteristics (for additional discussion see Error Resiliency, Chapter 4).

Power efficiency. In many cases, power will represent the limiting resource for both sensing and communications within a sensor network. It is recommended that sensor designs minimize power consumption in all cases and, in many cases, include the capability at the sensor level to initiate, accept, and survive partial or complete power down periods. For many sensors, this requirement may involve automation of normal warm up and initialization procedures. At the same time, sensor networks need to influence the development of and make use of stand-alone, benign power sources such as solar, wind, or fuel cell technologies.

Interactivity. In stand-alone mode, a sensor may retain little or no housekeeping information and rely entirely on user interaction for all operational parameters. In an array, that same sensor needs to provide interactivity with the network to allow flexible operations, including



remote setting of operation parameters such as observation frequency, signal gain, calibration frequency, etc. and transmission of appropriate operational and housekeeping metadata to the network. Each sensor should provide a self-description capability to the network, either through providing a full set of descriptors from the sensor, or through self-identification of the sensor according to preset sensor type descriptions held at a higher node in the array.

Affordability. In some experimental designs, the sensor elements of an array may meet observational requirements based more on their abundance and less on the precision and accuracy of individual sensors. In other experiments or settings, the nodes of an array may need to have extreme precision and accuracy to resolve or detect fine-scale, subtle, or transient features, signals, or events. The former application might allow use of relatively low cost sensors, while the latter application may require the use of relatively expensive sensors. In all cases, the affordability of a sensor will depend on measurement requirements, unit sensor costs, and sensor durability and reliability. In general, it is recommended that technologists evaluate or design sensors for array applications with low unit cost as a strong design parameter.

Environmental impact. A deployment of thousands of sensors as part of tens or hundreds of arrays, along with hardware, often batteries, and in some cases reagents at each node, could itself represent a form of environmental contamination. In addition, some active sensor technology emissions could be harmful or disruptive to nearby or even distant organisms. These factors need to be considered when designing sensors, and the possible environmental impacts of deploying large arrays of sensors needs to be examined in more detail.

2.2 Challenges and Solutions

Sensor arrays present new challenges for the design of sensor components.

2.2.1 Sensor Requirements from Array Networks

In the conceptual model of one or more sensor units connected to a node that provides power and communication infrastructure, the sensor units will require certain amounts of information from the array through the local nodes. At the design stage, sensor technologists need descriptions of the power, data, and communication interfaces of the host nodes. Ideally, in the modern electronics design environment, a sensor designer should have access to a node simulator or node interface

standard in order to test connectivity, compatibility, and throughput. In deployment, sensors will need timing and synchronization signals from the array nodes. In some scenarios, a sensor will also acquire position information from another (position) sensor at the same node. In other cases, however, a local sensor may require position information gained through network calculation processes. Network software should be structured to promote reusability of sensor device drivers and controllers across sensor types and generations. Moreover, the software should be structured to support multiplexing of multiple applications that might make use of a sensor output simultaneously *in situ*.

2.2.2 Network Requirements from Sensors

Scalability. Since sensor network topologies are dynamic and have heterogeneous sensor density, sensor design and configuration should promote ease of maintenance at all deployment densities. To reduce sensor maintenance for long term deployments, competitive sensor designs must provide both power efficiency (at a minimum) and, when possible, power harvesting from the environment. Power harvesting methods should avoid a large environmental impact, either in physical or chemical profile. Per-sensor maintenance should also be reduced to enable dense deployments, and sensor design must give high priority to the mitigation of degradation, contamination, and biofouling—both by the environment and to the environment.

Interoperability. Since sensor networks are envisioned to support a wide range of sensor types, data aggregation and profiling will be significantly eased by a common sensor network interface. This interface should include hardware, protocol, and data formats, as well as software structures that promote reusability. Sensor design should account for interoperability in sensor networks. While there are no common interfaces at this time and the need for standardization is still debated, several prototypical interfaces are emerging (e.g., Motes, TinyOS, and TinyDB; see Box 3).

Box 3. Research Experience in Self-organizing Networks: Motes, TinyOS, and TinyDB

Self-organizing wireless-sensor networks, a realization of the Pentagon's "smart-dust" concept, have reached the prototype stage worldwide. The smart sensors, or Motes, were created by the University of California at Berkeley and Intel, and are being tested worldwide today.

"At this stage, there are over 100 groups around the world that are using the combination of our open-source Motes with the TinyOS [operating system] and TinyDB [database]," said Berkeley professor David Culler, who is also director of Intel Research's "lablette" in Berkeley.

Researchers at the Defense Advanced Research Projects Agency (Darpa) proposed the smart-dust concept four years ago. The idea was to sprinkle thousands of tiny wireless sensors on a battlefield to monitor enemy movements without alerting the enemy to their presence. By self-organizing into a sensor network, smart dust would filter raw data for relevance before relaying only the important findings to central command.

The prototype Motes consist of an application-specific sensor array board married to a generic wireless controller board, both in a hermetically sealed enclosure. Once the design has matured, single-chip realizations will begin to downsize the wireless sensors to a volume less than a cubic millimeter. To facilitate the self-organizing of Motes into a sensor network, the researchers created TinyOS and TinyDB as well as a host of Tiny applications and a simulator[1].

References

[1] C. R. Johnson. Companies test prototype wireless-sensor nets, EE Times. January, 29, 2003. <http://www.eetimes.com/>

In addition to a common hardware and firmware sensor network interface, interoperability also requires a common framework for communication to and from the sensor. The great diversity of sensing modalities will entail extreme challenges in data integration and visualization without a carefully designed ontology to capture

the semantics of sensor interrogation and control. While there is much work to be done, several current NSF efforts capture this vision (Box 4).

The combined processing of data from multiple sensors with different observation positions and modalities

Box 4. GEON—The Geosciences Network

The Geosciences Network represents a coalition of IT and Earth Science researchers that has been formed in response to the pressing need in the geosciences to interlink and share multidisciplinary data sets to understand the complex dynamics of Earth systems. The need to manage the vast amounts of Earth science data was recognized through NSF-sponsored meetings, which gave birth to the Geoinformatics initiative [1]. The creation of GEON will provide the critical initial infrastructure necessary to facilitate Geoinformatics in support of a number of geoscience research initiatives, such as the EarthScope initiative.

Creating the GEON cyberinfrastructure to integrate, analyze, and model 4D data poses fundamental IT research challenges due to the extreme heterogeneity of geoscience data formats, storage and computing systems and, most importantly, the ubiquity of "hidden semantics" and differing conventions, terminologies, and ontological frameworks across disciplines. GEON IT research focuses on modeling, indexing, semantic mediation, and visualization of multi-scale 4D data, and creation of a prototype GEON Grid, to provide the geoscience community an "IT head start" in facing the research challenges posed by understanding the complex dynamics of Earth systems. An important contribution will be embarking on the definition of a Unified Geosciences Language System (UGLS), to enable semantic interoperability. The GEON Grid will leverage our experience in the National Partnership for Advanced Computational Infrastructure (NPACI) program, and the experience that will be gained in the recently awarded TeraGrid Distributed Terascale Facility. We will create a portal to provide access to the GEON environment, which will include advanced query interfaces to distributed, semantically-integrated databases, Web-enabled access to shared tools, and seamless access to distributed computational, storage, and visualization resources and data archives.

References

[1] GEON project summary: <http://www.geongrid.org/docs/summary.pdf>



requires each individual sensor to accurately characterize the sensing uncertainty of all data. In addition, the sampling time and the sensor location associated with data from each sensor are also needed for data analysis. If GPS signals are available in the observation environment, GPS is an effective technology to provide location and time information to individual sensors. Otherwise, the sensor network must provide alternative time synchronization and localization services to individual sensors.

Self description. Large-scale, dense deployment of sensor nodes will require a proportionately large investment in network configuration and maintenance, unless future sensor designs are able to incorporate self description, using the ontological framework described above. While self description includes identification and sensing attributes, the concept embraces much more. Self description is a task shared by both the sensor network interface and the sensor itself. For example, autocalibration among a population of nominally identical sensors will be an important byproduct of self description, and will ease the configuration of dense sensor deployments. As the deployment ages, an automated method of sensor self-inspection is a mandatory component of self description. Self-verification would answer such important questions as the residual lifetime of the sensor. Self description will also account for other sensor state information, such as the distance to other nodes and position estimation, or ranging.

2.3 Recommendations

In relation to the development of **new sponsored programs** in sensing networks, workshop participants made three recommendations.

- Sponsored programs should strongly encourage multi-institution investigations, especially coupling national research laboratories with universities. The national laboratories have a rich tradition in the development of advanced sensing capabilities, and they also have the capacity to produce devices at a scale needed for experimental work. Academic research efforts can benefit immensely from this cross-fertilization, while the national labs will benefit from dual use, an enlarged application space, and user feedback.
- It is also recommended that sponsoring agencies encourage university collaboration with industry through Small Business Innovation Research (SBIR) Programs, etc. Significant redundancy occurs in the configuration of different sensor networks. The goal of a common sensor network interface is only achiev-

able in industry (before market size is known) through government partnership.

- Cross-disciplinary collaboration is to be encouraged in sensor networks. These efforts should involve both sensor designers and end-user disciplines. This “culture-sharing” will provide greater consensus, application validation, and far faster progress on sensor durability, calibration, cost, sensitivity, etc.

Regarding the **design of sensors** for deployment in sensor networks, workshop participants made three recommendations.

- Technologists should reassess or modify current sensor design, taking into account the unique sensor requirements for sustained integrity, performance, and interactivity in long-term deployments.
- In all cases, hardware components of sensor arrays need to meet the highest standards for generating minimal environmental impacts.
- Sensor designs need to minimize power consumption in all cases, and in many cases should include the ability at the sensor level to initiate, accept, and survive partial or complete power down periods.

Box 5. SpecNet—Spectral Network

SpecNet (Spectral Network) is a network of sites that combine optical sampling with photosynthesis and respiration measurements, with the goal of improving our understanding of surface-atmosphere carbon and water vapor fluxes. An improved understanding of these fluxes is critical if we are to understand the biological controls on the Earth's changing carbon budget and the influence of surface properties on weather and climate.

SpecNet sites tend to be ones existing within FLUXNET, an international network of biosphere-atmosphere flux sampling sites that typically combine flux and optical sampling with a range of other sampling methods. SpecNet optical sampling incorporates a vast range of instruments, sampling protocols, and data formats. Examples of SpecNet optical sampling include spectral reflectance measurements and surface temperature measurements. In principle, this optical sampling is identical to “remote sensing” except that it is typically conducted from low-altitude field platforms (e.g., low flying aircraft, drones, mobile carts, or towers) rather than from satellites or high-altitude aircraft. Typically, fluxes are measured with eddy covariance towers, but can also be measured from aircraft or from whole-ecosystem chamber measurements (see Figure 1). Networked sensing exists at a few SpecNet sites, and offers considerable opportunity for improving our understanding of ecosystem flux and optical properties. A key SpecNet challenge is the integration of data from different domains (spatial and temporal) across a range of scales (individual organ to entire Earth) from disparate instruments and investigators. By matching optical and flux sampling in time and space, SpecNet is attempting to simplify this integration challenge.

Key technical goals of SpecNet include the standardization of field sampling instruments and field sampling and data storage and processing protocols. Unlike other Earth System Science efforts (e.g., FLUXNET and NASA's EOSDIS program), such standardization has not yet been applied to field optical sampling, and is essential if we are to conduct cross-ecosystem analyses. Such cross-site studies, and a suitable ecoinformatics database, are needed if we are to develop a broad understanding of physical and biological factors controlling surface-atmosphere fluxes. The development of such a database is in its early stages, and is largely an unfunded effort at this time.

Current SpecNet efforts are addressing factors that control flux rates, including physical factors (e.g., temperature, moisture, radiation, and topography) and biological factors (e.g., biomass or leaf area index, and species composition or functional type). Early results have revealed that contrasting ecosystems (e.g., shrubland and arctic ecosystems) are controlled by contrasting sets of environmental factors. At the same time, common limitations to carbon flux are emerging across a surprisingly wide range of ecosystems. For example, water limitations clearly restrict photosynthetic carbon uptake in both chaparral and arctic ecosystems.

Besides providing directly useful knowledge of the key flux processes for individual ecosystems, SpecNet is also providing a means of testing biospheric products emerging from current satellite sensors (e.g., MODIS). Early SpecNet results are revealing a number of potential problems with satellite sensors, which could lead to a redesign of some of the key algorithms used in processing satellite data for ecological purposes. Similar opportunities exist for developing and testing sensor networks within the diverse array of SpecNet sites.

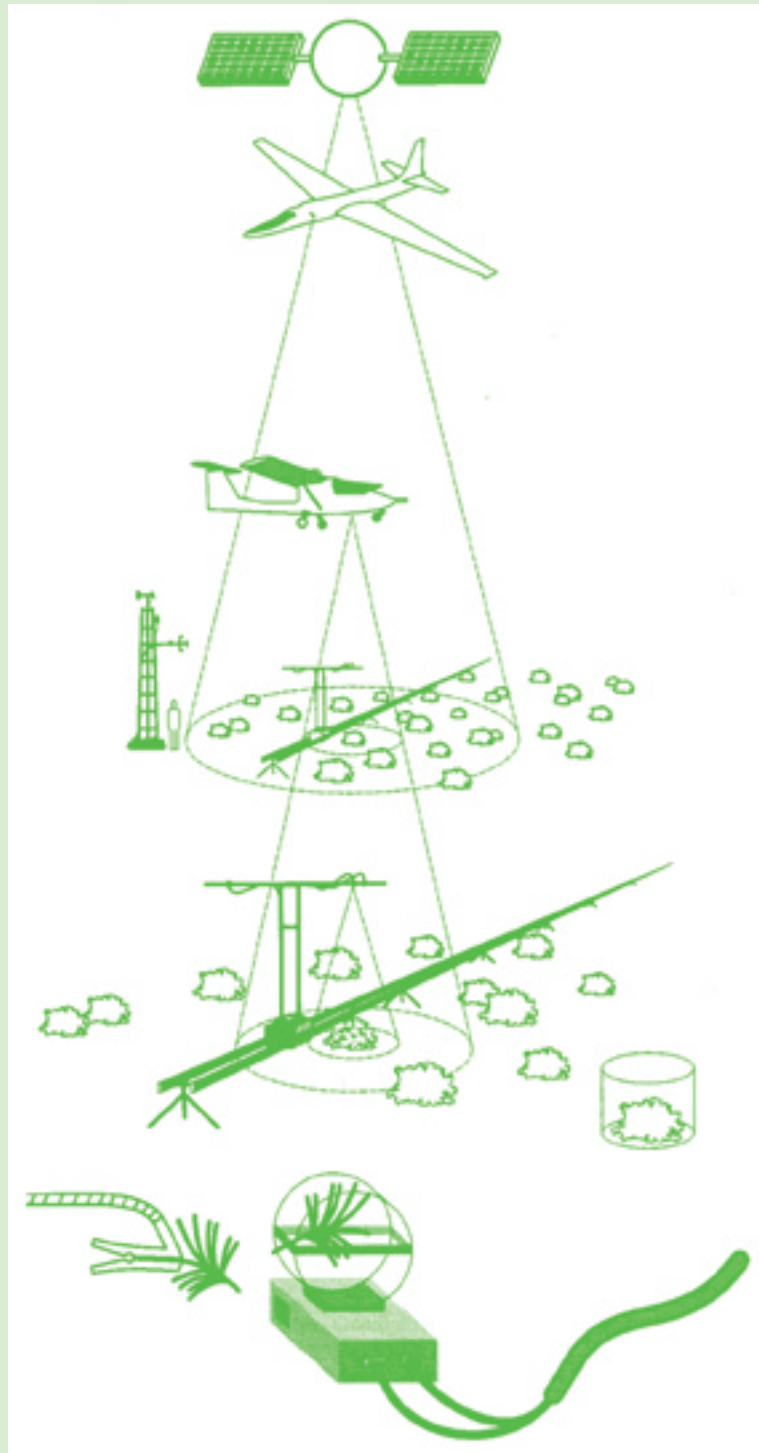


Figure 1. Illustration of multiple sampling methods used for the simultaneous flux and optical sampling within SpecNet. Sampling methods range from satellite and high-altitude aircraft (top) to low altitude aircraft, flux towers, and automated trams (middle) and chamber fluxes and leaf optical properties (bottom). SpecNet particularly emphasizes mid-range sampling methods illustrated in the middle of the figure, where the spatial scales of optical and flux sampling can be more readily matched. Finer scale methods (bottom) are used to interpret mid-range measurements, and large-scale measurements (top) provide a larger context.



Chapter 3. Deployed Sensor Arrays

The previous chapter addressed the properties of *individual* sensors. We now consider the issues that arise when incorporating individual sensors into a sensor *array*. The difference between the sensors and the array is one of *scale*—evident in both capabilities and challenges. For example, large geographic scale can give an array a far broader view of the environment than an individual sensor [1,2]. An array can also be more robust to failures by scaling up its density, giving multiple sensors a redundant view of the same phenomenon.

In the following sections, we examine sensor array issues in detail. In particular, we address the first question: Why do scientists need sensor arrays rather than individual sensors? Second, we identify the challenges in creating sensor arrays. Finally, we suggest where future programs should be targeted so as to most effectively advance the field.

3.1 Introduction

Evolving scientific needs for new types of observation are the driving force behind the creation of sensor arrays. There are many phenomena that cannot be observed by individual sensors, and are only observable by a sensor *array* of sufficient scale.

The advantages of scale apply across many dimensions. Perhaps the most important is geographic scale: an array covers a much larger area than a single sensor. This property is crucial for phenomena that both cannot be observed from far away and are distributed over a large geographic area.

There are phenomena that have only one of these properties, and are well served by individual sensors rather than arrays. For example, storm clouds can cover a large area, but are also observable from far away; meteorologists can use a few centralized sensors such as Doppler radar or satellite imagery for these observations. Conversely, the temperature or rainfall at a single point can only be measured locally—it is unobservable from far away, but only requires a single sensor.

3.1.1 Observing the Previously Unobservable

While remote sensing has been highly effective in some contexts, it has significant limitations in others. For example, some sensors (e.g., imagers) require line-of-sight; this may not be available in some environments under observation (e.g., belowground ecosystems, dense forests). In

some cases, the signal of interest may not propagate very far through the environment—for example, the high-frequency components of seismic waves do not propagate well through the earth.

Observation of the environment with a single sensor is not feasible in cases where remote sensing is impossible *and* the phenomenon is geographically widely distributed. The need for sensor arrays is thus born from the growing needs of many scientists to observe phenomena that are *both* distributed *and* difficult to sense remotely. Such conditions arise particularly in two areas:

Observation of distributed environmental/ecosystem processes. For example, consider measurement of the carbon flux of ecosystems. There is no remote sensor in existence today that can fully characterize carbon flux from soils or aquatic systems, nor can such a measurement, taken only at one location, be accurately extrapolated across an ecosystem. Only with the aid of fielded arrays of *in situ* sensors can carbon measurements be made across landscapes and earth's ecosystems with the accuracy needed by researchers, policymakers, and ecosystem managers. Similar needs for sensor arrays apply in a range of other phenomena: the spatially and temporally dense measurements needed to understand terrestrial, oceanic, and atmospheric material cycle interactions; ecosystem and geological processes including seismic activity; complex climate phenomena such as El Niño; and the movement of nutrients and pollutants through groundwater and the atmosphere, and between ecosystems.

Sensing of organisms over a range of spatial and temporal scales—for example, to track their movement within and between ecosystems—can only be accomplished with the aid of *in situ* sensor arrays distributed across landscapes. The spatial resolution (i.e., pixel size) of remotely collected data may be too coarse to identify even individual tree species, let alone small animals under the canopy, fungal organisms living in the soil, or bacterial spores blowing in the wind. The use of distributed sensor arrays taking close-in measurements at many locations, in combination with other common technologies including radar, opens the door for researchers to gather the vector information that is necessary for tracing the movement of organisms, including migratory birds, endangered species, invasive species, game animals, and wind-borne pathogens through space and time. Such arrays will also facilitate a variety of organismal studies, revealing previously unobservable phenomena

associated with animal behavior, plant phenology, and below-ground ecosystems.

3.1.2 Validation, Cross-checking, and Data Fusion

The previous section describes the observation of new phenomena made possible by sensing over a large *spatial* scale. Scientists also have research goals that can leverage a sensor array's *density* scale. That is, as an example of "the whole being greater than the sum of the parts," when many sensors make observations of the same phenomenon at the same time across a region, the data can be integrated in ways that increase its value—making it more reliable, less ambiguous, finer resolution, or correlated with other aspects of the environment.

The simplest examples of validation come from placing multiple, redundant sensors in view of the same phenomenon. At a minimum, a basic scheme can detect failures by building a consensus across sensors before reporting a value. More complex schemes can use multiple views to provide automatic calibration. If one sensor is significantly less prone to error than others (e.g., a recently repaired or a higher quality sensor), it can be used to calibrate other sensors in the array [3].

A more fundamental improvement in the quality of data can be achieved by using a set of *heterogeneous* sensors—that is, sensors in the same area that have a diversity of modalities, dynamic ranges, failure modes, and so forth. Biological systems are so complex that one must compare and integrate data from multiple sources to validate conclusions and gain a deeper understanding of the underlying phenomena. The need for multi-modal sensing also drives the need for sensor arrays over individual sensors.

3.2 Challenges and Solutions

The vision for large-scale sensor arrays is compelling, but significant challenges must be addressed before such arrays will be practical. Scale drives many of the challenges, but at the same time creates new opportunities.

3.2.1 The Harshness of the Real World

Sensors deployed in the real world are subject to harsh and sometimes unpredictable conditions. Hardware must operate over a wide range of temperatures, be resilient to moisture, and have a reliable source of power. In many cases, a sensor's operation may require careful physical orientation—for example, a camera may need

to be looking directly at a particular bird nest, a rainfall sensor can't be covered by leaves, and a chemical sensor might need to be immersed in a stream. Devices can be disturbed by weather (wind, rain); animals (eating cables, moving sensors); or people (curiosity, vandalism) [4].

While these issues plague both arrays and individual sensors, in an array the problem is exacerbated by scale: a larger array implies less attention (e.g., maintenance) available for any single sensor, and sensors can experience a wider diversity of failure modes. However, while these factors make the array more prone to failure, at the same time its scale provides redundancy so that failures have less impact. For example, several imagers can be aimed at the same area, so that observation continues even if a single camera fails or moves. Thus, a challenge is to design a new generation of sensors that individually have unprecedented *autonomy*, and collectively have a *high tolerance for individual failure*.

There are a number of technological challenges that must be overcome to realize this vision for easy deployment of large-scale networks, and the use of redundancy to seamlessly mask faults. These challenges include:

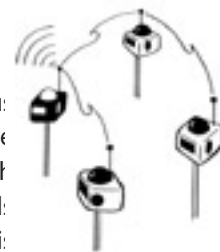
Self-configuration. Sensors in an array must automatically determine their role and configure themselves appropriately. Unlike desktop computers, which enjoy one-on-one attention, a single user may be responsible for thousands of sensors, making manual configuration infeasible [5,6,7].

Graceful degradation. Node failures should not simply bring the array to a halt. As more nodes fail, the array needs to continue to provide the best service possible for the available resources.

Sensor and component robustness. Despite automatic reconfiguration and graceful degradation, eventually a sufficient number of failures will prevent the array from fulfilling its intended purpose. Thus, if the array is to be long-lived (months or years), resiliency features do not eliminate the need to build node hardware that is sufficiently robust and rugged to operate in harsh environments over long periods.

Backward compatibility. In long-lived, large-scale deployments, several generations of hardware and software will be expected to interoperate, as new sensors are built and added to existing networks.

Taskability/programmability. Node software is often highly domain and task specific, but the deployed array in many cases will outlive the task for which it was initially intended. Thus, it is desirable that nodes be reprogrammable or re-taskable *in situ*.



3.2.2 The Diversity of Spatial and Temporal Scales

The many kinds of observations in which scientists are interested range along a spectrum of scales, from nano-seconds to centuries, from molecular to global. At one end of the continuum might be seismology, in which high frequency continuous measurements are recorded on large physical-scale devices with low spatial resolution. At the other end may be instrumentation of ecosystem processes such as respiration; this has far lower temporal frequency but calls for very small form factor devices placed at high spatial density. In the middle of the continuum might be acoustic arrays, which record high frequency signals intermittently, at moderate spatial resolutions and form factors.

Systems dominated by high temporal resolution face challenges primarily associated with very high data rates, while those dominated by spatial resolution are most challenged by the resource limitations associated with small form factor devices.

Even within an individual study, observations may be required over a wide range of scales. For example, in the study of a global carbon cycle model, small-scale *in situ* measurements of carbon flux from soils, oceans, and freshwater ecosystems must be integrated with larger scale remote observations of ecosystem, biome, continental, and oceanic aboveground vegetation and net primary production (NPP). This is a challenge due not only to the scaling issues involved and the widely disparate data sets that must be integrated, but also because a diversity of sensor arrays will be needed for each of the particular ecosystem components being studied, whether belowground, aboveground, or aquatic components. In the case of episodic events (forest fires, floods, volcanic eruptions, earthquakes, hurricanes), drastic changes may occur in relatively short periods of time. Unless networks of distributed sensors are in place across broad spatial scales to measure the phenomena before, during, and after, much of the material fluxes and transformations that occur will be lost. Such data gaps hinder current ecosystem and global modeling efforts, and impede informed, science-based decision-making.

3.2.3 The Costs of an Array

As with any technology, a sensor array must have a low enough cost to be practical—not only in terms of hardware cost but also in terms of personnel time for development, deployment, and maintenance.

The hardware cost itself is perhaps the most obvious issue: some specialized sensors can be quite expensive (especially those with both large dynamic range and high resolution). In some cases, buying an array of hundred or thousands of such sensors is simply not feasible. This workshop is partly a call to develop lower-cost sensors, and also a challenge to use tiered architectures—that is, increasing the value of data from large numbers of lower-cost sensors by supplementing this with information from a few higher-cost, higher-value sensors [1].

The cost of maintenance also has the potential to scale with the number of nodes in the array. In an array with hundreds or thousands of nodes, it is not practical to produce a design in which every node requires personal attention from a human. The cost to replace failed components will also scale with the array size. These challenges must be met through software advances such as the self-configuration, and the graceful degradation architecture described in the previous section.

Software development cost is another important issue. On the surface, it would seem to be the easiest cost component to minimize, as it does not directly scale up with the size of the array. And unlike hardware, some software development can also be shared among collaborators. However, sensor arrays are often highly domain-specific and even task-specific (e.g., algorithms for in-network data reduction, summarization, or aggregation, which are often needed to meet the channel capacity or energy requirements as the network scales) [8,9]. This introduces another, subtler problem, which requires cross-discipline collaboration. While it is the scientist—not the technologist—who understands the domain well enough to know which data can be thrown away or summarized and which data must be transmitted with perfect fidelity, it is the technologist who must express those policies as program code that runs inside the network. This cross-discipline problem has been seen in other domains (e.g., physics and molecular simulations).

Finally, the environmental impact of sensor arrays needs to be considered. We will soon reach a point where sensors are so cheap and deployed in such great numbers that it will be more practical to manufacture new arrays than to retrieve them from the field. This is an important challenge which must be met in order to avoid contaminating the environment that this research seeks to protect—perhaps through advances in biodegradable materials, perhaps through careful deployment and retrieval policies.

3.2.4 Multidisciplinary Collaboration

The previous section describes a challenge in software development for sensor arrays: the application domain knowledge and the ability to express the resulting policy come from different disciplines. This is one aspect of a more general challenge: sensor arrays cut across many disciplines, none of which can operate in a vacuum. Without cross-disciplinary communication and collaboration, the value of array sensing can be significantly reduced.

For example, while technologists are often driven to attack the most technically interesting problems, it is also important that they recast their questions to produce answers of practical value to scientists. Similarly, scientists need to begin to recast their biological or ecological questions so as to leverage the new types of observations made possible with emerging sensor arrays, thus giving technologists tangible goals rather than leaving them to operate in a vacuum. Only through such collaboration can technology be created that ultimately serves a purpose for both scientists and society (non-technologists).

3.3 Recommendations

Deployed sensor arrays can offer radically new perspectives on key scientific challenges. As described above, fundamental to this capability are *system scaling* (the ability to deploy and exploit very large numbers of observations over time and space) and *heterogeneity* (the ability to deploy arrays in support of a wide range of scientific questions across a wide range of physical environments and phenomena). While the scientific and engineering communities agree on the importance, potential, and challenges associated with deployed sensor arrays, it is only through programmatic action that the benefits of this technology can be realized. In particular, we propose the following four programmatic recommendations:

3.3.1 Enable System Scalability Necessary to Realize the Vision

Funding and coordination mechanisms must be introduced to achieve system scalability in the near term. Without appropriate mechanisms/interventions, it will not be feasible to design, implement, deploy, operate, and use scalable sensor arrays in the timeframe needed to understand and address critical environmental issues. We recommend the following specific interventions:

- Coordination of research funding and communities horizontally across science domains to achieve economy of scale in hardware and software reuse.

- Coordination of research funding and communities vertically across science and information technology disciplines to realize the scaling possibilities offered by existing techniques.
- Undertake long-term research in information technology to support the ultimate vision of broad spatial scales, hyper-density, and hyper-heterogeneity in deployed arrays.

3.3.2 Integration of Expertise to Define New Directions and Priorities

It has become widely recognized that many of today's scientific and engineering advances are being realized through multidisciplinary interactions. Scientists will be able to exploit sensor array technology if and only if they work with information technology researchers to *understand* the emerging capabilities of distributed sensor arrays, and participate intimately in *defining* key system characteristics at the high and low level (i.e., component and system characteristics described in other chapters of this report). In particular, we recommend interventions to ensure the following:

- Scientists should be supported in the creation of *next generation science questions* that take into account the emerging technological capabilities of sensor arrays. Only in this way will science go beyond simply more data points more cheaply to truly revealing previously unobserved phenomena. In addition, technologists and scientists must work together on an ongoing basis to set and refine priorities for technological developments that address key scientific needs.
- More general information technology research related to scalable sensor arrays is essential to achieve the leaps in capability (e.g., micro-scale sensors and platforms, very low power devices, real-time data availability, self-configuring and autonomous massively distributed algorithms, *in situ* repairability) needed for hyper-scalability and heterogeneity.
- In both science and engineering, new educational training programs are needed to foster scientists well-versed in the technical capabilities of sensor arrays and the appropriate methodologies for their use. Similarly, information technology students can benefit tremendously from the inspiration of the challenges and concrete requirements of their science colleagues.
- Outreach to decision-makers (i.e., users of the science) must be supported. Deployed sensor arrays are of tremendous relevance to resource management policy makers and practitioners, both with respect to their need for the scientific understanding that will be

enabled by this capability, and in their own direct use of deployed arrays for monitoring and assessment.

These integrating actions must be reflected in funding mechanisms, as well as in key coordinating mechanisms such as workshops and science and technology centers.

3.3.3 Process for Moving from Capability to Deployability

Some sensor arrays will be deployed across a scattering of applications and locales without any intervention. However, sensor array technology is unlikely to achieve true scalability without interventions that facilitate the development of application-specific arrays whose system components (both hardware and software, both node and network-wide) are designed, built, and verified to be usable across a range of applications and target domains. To achieve this, we suggest the following interventions:

- Funding and activities should be coordinated to pursue a series of diverse yet specific (well-defined application) pilot deployments from which to generalize.
- Systematic methods for moving from pilot to deployment target (terrain and scale) should be developed through research and experimentation.
- Incentives for and ease of contributing to open source tools and testbeds are essential to shift the community toward development of broadly used and reused system components. These include hardware, software, communication protocols, and tool development, as well as complete reference implementations.
- There must be focused support for *staffing* to carry out the development of reusable systems, and to develop and rigorously characterize these systems in terms of metrics that are meaningful to science applications and tractable to technologists. Faculty and graduate student support alone are not sufficient for creating such comprehensive, cross-cutting infrastructure.

3.3.4 Sustain Long-term Deployments

Finally, as the technological and scientific advances are made that are needed to deploy sensor arrays in the service of science, we must ensure that we sustain these critical national resources. To this end we suggest the following interventions:

- Funding models are required that recognize the importance of staffing costs for stewardship and management.

- Up front and ongoing coordination is essential for contributing to integrated heterogeneous facilities and data sets.
- Review processes are needed that will keep facilities alive, evolving, and non-obsolete.



3.4 References

- [1] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao. Habitat monitoring: application driver for wireless communications technology. *Proceedings of the 2001 ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean*. San Jose, Costa Rica, 3-5 April, 2001.
- [2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, J. Anderson. Wireless sensor networks for habitat monitoring. *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications*. Atlanta, GA, September 28, 2002.
- [3] K. Whitehouse and D. Culler. Calibration as parameter estimation in sensor networks. *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications*. Atlanta GA, September 2002.
- [4] J. Polastre. Design and implementation of wireless sensor networks for habitat monitoring. Master's Thesis, University of California at Berkeley, Spring 2003.
- [5] N. Bulusu, J. Heidemann and D. Estrin. Adaptive beacon placement. *Proceedings of the Twenty First International Conference on Distributed Computing Systems (ICDCS-21)*. Phoenix, Arizona, USA, April 2001.
- [6] A. Cerpa and D. Estrin. ASCENT: Adaptive Self-Configuring Sensor Networks Topologies. *Proceedings of the Twenty First International Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002)*. New York, NY, June 23-27, 2002.
- [7] L. Girod, V. Bychkovskiy, J. Elson, and D. Estrin. Locating tiny sensors in time and space: a case study. *Proceedings of the International Conference on Computer Design (ICCD 2002)*. Freiburg, Germany. September 16-18, 2002.
- [8] J. Elson, S. Bien, N. Busek, V. Bychkovskiy, A. Cerpa, D. Ganesan, L. Girod, B. Greenstein, T. Schoellhammer, T. Stathopoulos, and D. Estrin, EmStar: An environment for developing wireless embedded systems software. *CENS Technical Report 0009*. March 24, 2003.
- [9] J. Gao. Energy efficient routing for wireless sensor networks. PhD Dissertation, University of California at Los Angeles, August 2000.

Box 6. Embedded Networked Sensing

Embedded Networked Sensing (ENS), the deployment of wirelessly networked sensors throughout an ecosystem, will rapidly increase our understanding of natural and anthropogenic phenomena on Earth [1]. A globally significant application for ENS is to monitor the spatial and temporal dynamics of biocomplexity in its environmental, ecological, and cultural conditions, and the interactions between these dynamic processes in natural and human landscapes. The vision for habitat sensing applications at the Center for Embedded Networked Sensing (CENS; <http://cens.ucla.edu>), a NSF Science and Technology Center established in 2002, is the development of robust tools that can operate remotely in uncontrolled natural and agricultural settings, and that can capture and integrate data across a wide range of ecological scales.

CENS is currently developing and deploying standardized, inexpensive, and lightweight but durable networked sensing systems to collect data in a variety of ecosystems across a range of temporal, topographic and ecological gradients. Key microclimate sensor systems are currently being coupled with actuated video cameras to document the visual status of animals and plants in response to short term microclimate changes, and for proximity and change detection.

The instrument array is a hierarchical, wireless remote data logging system powered by a combination of solar panels and deep cycle batteries. It has the capability to log and transmit core small-scale environmental data to a central facility, as well as to accommodate analog and digital data inputs useful in controlling micro-video cameras, and proximity sensors. Long-lived unattended operation requires that we imbue the distributed system with a form of distributed intelligent operation such that raw time series data are transformed through local processing from raw data into information of interest. Such in-network processing is critical to the scalability and longevity of these systems, as well as to the incorporation of higher capability sensing and sampling that cannot be densely deployed



Figure 1. A hypothetical Networked Infomechanical System (NIMS) in a forest ecosystem.

in situ [2,3]. Ultimately these systems will embody a complete ecology of their own, from small static nodes with limited sensing range and capabilities, to higher end nodes that aggregate, and process at higher levels, to the autonomously robotic elements that will enable true spatial and technical diversity (see Figure 1). This latter technology, termed NIMS (Networked Infomechanical Systems http://www.cens.ucla.edu/censweb/Project_Page_WEB_UPDATES/NIMS/) will lead not only to implementation of mobile sensor nodes, but also the ability to remotely collect air and/or water samples. At CENS, engineers working with biologists are just beginning to tap the potential that ENS systems offer for understanding our world.

References

[1] CSTB Committee on Networked Systems of Embedded Computers. *Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded* National Academy Press, 2001. http://www7.nationalacademies.org/cstb/pub_embedded.html

[2] D. Estrin, D. Culler, K. Pister, and G. Sukhatme. Connecting the physical world with pervasive networks. *IEEE Pervasive Computing*, 1(1): 59-69, 2002.

[3] G. Pottie and W. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5): 51-58, 2000.



Chapter 4. Error Resiliency

The distributed and redundant nature of sensor arrays offers both many more opportunities for errors to be introduced, as well as more opportunities for increased error resiliency. However, realizing these benefits will require integration of domain-specific, sensor-specific, and array-specific knowledge. An important challenge will be to extract common, reusable components and to spur the development of shared infrastructure and tools.

4.1 Introduction

In the context of large-scale, heterogeneous sensor networks, we define *error resiliency* as the ability of a sensor network to recognize and respond to sources of error via processes of error containment, management, and mitigation. The goals of an error resilient sensor network include:

- Providing the user with high quality information in the presence of most, if not all, forms of error and uncertainty (e.g., those appearing at both sensor and network level), under a variety of resource limitations and tradeoffs (e.g., power, bandwidth), while taking into account various user priorities (e.g., immediacy, data fidelity).
- Providing the user with measures of confidence or reliability that describe errors and uncertainty in the data set (both current and future, i.e., predictive).

When these goals are met, preferably in ways that are transparent to the user, the sensor network will provide and sustain the desired quality of service to scientists and the broader user community.

There can be many sources of error and uncertainty in a sensor network. Errors routinely occur at the sensor level due to inherent limitations such as quantization of analog data or thermal noise. External factors can also impact sensor operation in a harsh environment, and can, for instance, cause mis-calibration, hardware and software faults, and errors in time or space localization. The aggregation of all of these error sources has the potential to place the individual sensor outside its “accepted” operating point or specification, leading to errors in local decision-making, data measurement, data collection (e.g., summarization), and data processing (e.g., compression).

Network level errors can also greatly impact the scientific mission of a sensor network. These include errors due to noise, signal/packet fading or loss, mis-configuration, routing faults, or application failures at the network level (e.g., file transfer). Some of the key characteristics of sensor networks, such as episodic transmission of data within an energy-constrained network, can make debugging and error containment or analysis a significant challenge. Propagation and accumulation of uncertainty in a networked environment is yet another source of error that impacts the network’s performance and robustness. Aliasing and resolution errors can also occur as a result of a poorly deployed network.

It is important to note that error resiliency and error containment/management are phenomenon- and application/domain-specific. As such, solutions developed for one application may not work for another, or may be insufficient or even wasteful of resources. Here again, an interdisciplinary effort that includes a thorough understanding of the problem domain and error-handling requirements of the end-users is paramount, before error resiliency can be incorporated into a sensor network. Nonetheless, there are a number of general challenges and needs that are pervasive in nature, and will apply to a large class of sensor networks. We elaborate on these below.

Error resiliency can be divided into the resiliency of individual physical sensor components and the resiliency of synthetic sensors that are comprised of the behavior of a collection of sensors and processing.

4.1.1 Sensor and Network Components to Meet Performance Targets

Understanding sensor network component errors will increase our ability to quantify and manage the uncertainty associated with the data delivered by the sensor network. Scientific measurements are generally prefaced by instrument calibration against known standards, to yield information about the errors and precision of the instrument. Sensor network calibration methods will be similarly important. In addition, the overall sensor network must tolerate error or faults without failing [1]. This discussion of requirements is associated with learning how to calibrate the components of a sensor network, and does not address errors in network coverage, which are discussed below. Sensor network component errors

include those associated with sensors and networks, and are related to:

Sensor precision. The capacity of a sensor to reproduce the same signal under the same conditions is generally specified, but may change or degrade over time and under application conditions. In some cases, these errors are random, and data taken on the average are sufficiently precise. For example, a temperature monitoring network made up of thermistors with an individual precision of ± 0.1 °C, may report values reflecting this level of “noise.” In an application that includes observing diurnal temperature variation of tens of degrees, this level of error may be acceptable.

Sensor accuracy. Raw sensor signals are generally related to the observational information that they sense by physical relationships or models. These relationships are well-understood and highly accurate for some properties (e.g., temperature from a thermocouple response), but less so for others, particularly when the geometry of the sensor-medium interface is important, or when the medium must be disturbed by sensor deployment (e.g., soil moisture sensors).

Sensor locations in time and space. Self-locating or human-configured locations and time-synching errors will need to be addressed [2,7].

Loss of data or node activity (network fault tolerance). There is a need to adapt existing or develop new algorithms for coping with the loss of connections between sensors or individual nodes in a sensor network. In a remotely deployed network, for example, the loss of a few nodes may be acceptable if the network can dynamically adapt to this node loss [1].

As this range of categories implies, every sensor network is created with some inherent amount of component error, and this error is likely to grow as the network physically interacts with its environment.

4.1.2 “Synthetic Sensors” to Meet Performance Targets

Much of the promise of distributed sensor systems lies in their ability to synthesize data collected from a distributed set of sensors into a single logical result. For example, the values from a collection of distributed precipitation sensors can be combined to yield the overall precipitation for a watershed. In traditional systems, this type of synthesis is performed in a subsequent or post-processing data analysis step that processes data after it has been retrieved from the network of sensors. For sen-

sor networks, several trends are combining to make this type of solution less desirable:

- As the scale of sensor networks increases, the amount of data available in the network increases.
- The requirements for power savings continue to apply pressure to reduce data communications requirements for these systems.
- Emerging real-time in-network analysis capabilities provide opportunities to modulate the types of data collected with low latency (e.g., photograph moss after onset of rain).

To address these problems and opportunities, we propose the concept of a *synthetic sensor*, which is a system of devices, sensors, software, and protocols that collectively provide a logical interface to a result ready for scientific interpretation. This idea is in some ways an extension of preexisting sensor technology (e.g., temperature compensated sensors) to a much more complex distributed system environment. Synthetic sensors are similar to simple sensors in many ways:

Characterization. The error properties, domain, and range of a synthetic sensor must be characterized.

Calibration and validation. There must be a way to calibrate the sensor and validate its operation in a given environment.

Confidence estimation. The capability of reporting confidence information is important. In many ways, this ability is even more crucial for synthetic sensors due to the fact that their behavior, including range of valid operation, is more complex. Because of their redundancy, synthetic sensors will often have greater capacity than simple sensors to provide confidence estimates.

4.2 Challenges and Solutions

4.2.1 Challenges in Sensor and Network Components

Determining general calibration strategies. For large-scale sensor networks, manual, single-sensor calibration schemes do not work well. In addition to the obvious scaling issues, limited sensor access and complex environmental effects such as those discussed below may pose a still greater challenge to sensor network error management efforts.

Range of sensor types. As noted in Chapter 2, the potential range of physical, chemical, and biological sensors is large and growing. Being able to know the accuracy of these sensors in the context of their applications will



require extensive validation. Developing algorithms for quantifying, combining, and propagating errors across the spectrum of sensors is a major challenge that must be addressed.

Range of deployment applications. Providing confidence in specifying the location in dynamic (flowing/blowing) environments as well as the timing/synchronization of sensors, along with developing methods and algorithms for correcting these errors, are major challenges that need to be addressed.

Environmental effects. Environmental factors will influence sensor network accuracy and precision. These include temperature, water or air quality conditions, biological fouling, and other factors [3,4,5]. Developing and standardizing calibration procedures for this broad array of variables will require a significant effort.

4.2.2 Challenges in Synthetic Sensors

The cross-disciplinary nature of the problem. Before data analysis can be “pushed into the network,” a clear understanding of its operation is needed. In practice, the implementation of chosen analysis algorithms will likely require iteration between the end-user scientist who developed the analysis techniques and the technologists charged with implementing them. For example, the issues include:

- What data are needed where, and with what latency and reliability?
- What are the advantages of centralized versus localized analysis algorithms?
- How can analysis techniques culled from other fields be applied, e.g., a vision algorithm to detect birds in a nest?

Achieving generality despite the application-specific nature of each problem. Many of the data analysis techniques employed will necessarily be application-specific. In some instances, the structure, and hence some aspects of the distributed system, may also be application-specific. One of the more important challenges and opportunities will be to distinguish and articulate the general principles and techniques from the application-specific details of the first few applications.

Characterization challenges. In general, the characterization of a synthetic sensor is more difficult due to its increased complexity. For example, because it is composed of many separate components, characterizing the large number of possible failure modes is a significant

problem. The characterization of a synthetic sensor will need to take into account:

- **Faults in sensors and nodes versus redundancy.** In a large population of sensors and nodes, failure of individual components is a certainty. However, a synthetic sensor can potentially leverage redundancy to overcome these failures. Part of the sensor characterization must address how these failures and compensatory techniques affect the overall performance of the system, and how to factor this into the assessment of confidence.
- **Aggregation and lossy compression.** In order to meet operational requirements, a sensor network may need to make more efficient use of communication channels and energy by reducing the amount of data transmitted and processed. Thus, the results computed by a synthetic sensor may be based on summarized, aggregated, or compressed data, which in some cases have lower information content than the original. Part of sensor characterization must address how this data reduction affects the overall performance of the system.
- **Capability and applicability of algorithms.** The algorithms developed to analyze the data and to generate the synthetic sensor output must be characterized. A given algorithm may have regions of applicability that define the domain that the sensor can sense, as well as the range of valid outputs.
- **Coverage.** In order to produce valid data outputs, a synthetic sensor must have adequate coverage of the phenomenon being observed [6,7]. A synthetic sensor might be characterized in terms of its response to particular coverage scenarios. However, a detailed characterization may *not* be needed, as long as there are tools that help an end-user deploy the system with sufficient coverage.
- **Calibration and validation challenges.** Similar to simple sensors, synthetic sensors must be calibrated and validated before they can be used. Due to the increased complexity of a synthetic sensor, calibration and validation are much more difficult tasks. Although redundancy in the system may mean that the system can self-calibrate to some extent, it is likely that in many cases validation will have to be done by comparison to ground-truth data in an instrumented test.

Continued on page 30

Box 7. NSF CLEANER Initiative

CLEANER (Collaborative Large-scale Engineering Analysis Network for Environmental Research) is an integrated network of state-of-the-art user facilities to support cyberinfrastructure needs for engineering, research, and education on large-scale, cross-cutting issues-based environmental problems. It will provide researchers across the nation access to leading edge linked sensing networks, characterization tools, and data repositories and computational tools for integrated assessment modeling. Modeling would be a central component for analysis, knowledge synthesis and design of further experimentation. Specifically, the integrated models will allow both reductionist and multidisciplinary researchers to synthesize knowledge about diverse environmental settings and to readily identify knowledge gaps leading to improved theory. Collectively, CLEANER will provide the capabilities for near-real-time dynamic monitoring and analysis of parameters that are key to effective environmental management. Thus, CLEANER will be a cyberinfrastructure “test bed” as an engineering analysis network.

CLEANER will enable the development of integrated community models of anthropogenically-stressed large-scale environmental systems, such as the coastal margins and river and estuary systems. The CLEANER cyberinfrastructure will promote multidisciplinary research on adaptive environmental management and a testbed for engineering cyberinfrastructure investments. This will provide a focus for developing and/or defining:

- User needs and system architecture, software, hardware, technical support, and outreach and training for effectively addressing these needs;
- Innovative high-performance sensors;
- Configuring, siting, and operating integrated sensor networks;
- Advanced modeling capabilities;
- Collaborative tools;
- New tools and strategies for storing and accessing, manipulating, analyzing, integrating and visualizing diverse data sets;
- Common data handling protocols and standards; and integration of experimentation and simulation.



Figure1. CLEANER will provide the capabilities for dynamic monitoring and analysis that are key to understanding and management of complex environmental systems.



Figure2. Sensor-based autonomous robotic systems support dynamic monitoring and analysis of complex environmental systems.



Engineered sustainability of a coastal margin region provides a good illustration of a large-scale problem that cannot be addressed by conventional individual research projects. Such regions occur where fresh water sources reach coastal areas. As shown in Figure 3, the largest recent growth in United States population has occurred within approximately 100 miles of the coasts. These populations are dramatically affected by environmental issues in these rivers, estuaries, and coastal waters. Regions such as the Chesapeake Bay, the Hudson River and estuary, the Neuse River, the Mississippi delta, Corpus Christi Bay, and Santa Monica Bay, and the Colorado River basin are all examples where events involving contaminants, biological hazards, algal blooms, water shortage, and engineered interventions have affected large populations. An integrated and comprehensive knowledge about these systems acquired through the CLEANER program would support the evaluation of alternative engineered solutions and policies.

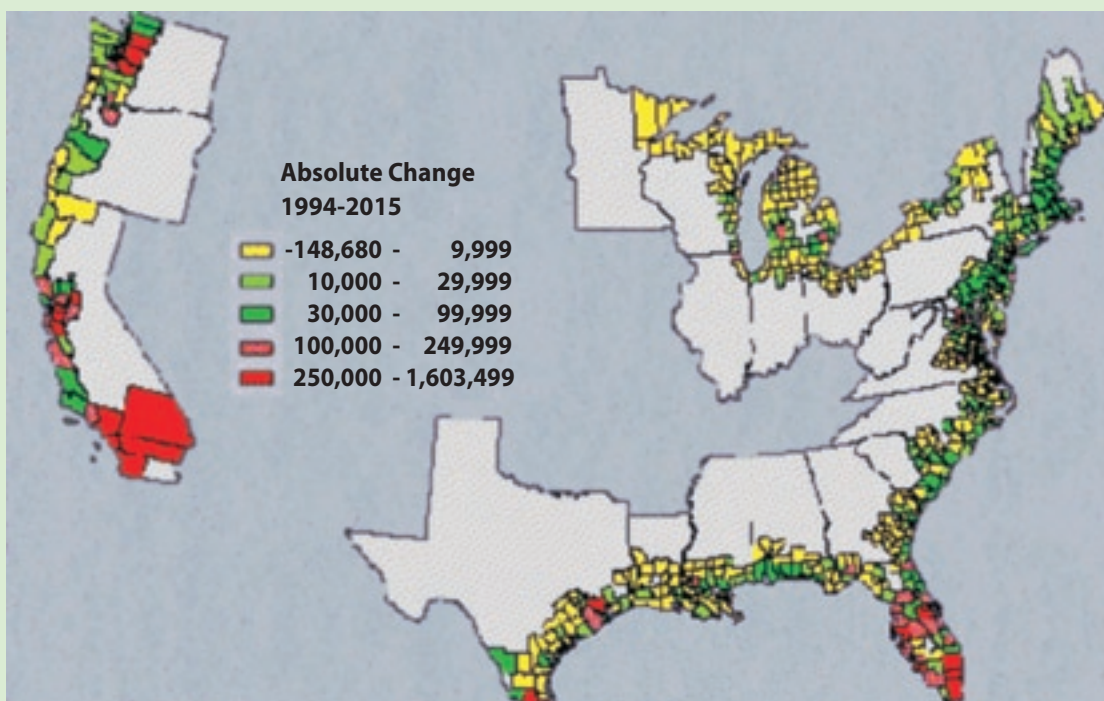


Figure 3. Projected U.S. population change from 1994-2015

4.2.3 Challenges in Adaptive Reconfiguration for Error Resiliency

Successful error containment and management depends significantly on the ability of a sensor network to adaptively reconfigure itself in response to a number of possible disturbances or events. Examples of these events include:

- Errors and faults.
- Environmental changes (e.g., changes in temperature, humidity, chemical composition and concentration).
- Security intrusions (see Chapter 5).
- Changes in service level (e.g., from low data rate to high data rate; from text or command data to images).

In the future, sensor networks are likely to be highly heterogeneous, sensing a diversity of parameters, performing a variety of *in situ* computations, operating with multiple communication protocols, and providing an array of services, from simple periodic samples to continuous audio and video. Furthermore, sensor networks will have to respond to an increasing range of unknown error events, faults, and sudden changes in their operating environment. Dynamic reconfiguration will be key to containing errors and maintaining a desired level of operation.

Examples of areas in which errors can affect sensor network performance, requiring an adaptive reconfiguration capability, include, but are not limited to:

Resource allocation. Limited resources must be allocated adaptively to achieve the level of service required for a given science application. These resources include system components and energy resources, as well as network resources such as bandwidth, latency, and link quality. Some typical mechanisms addressing these issues include duty-cycling, link admission, load balancing, rate limiting, congestion control, and route availability/adaptability. These mechanisms must be resilient to errors and losses in the control traffic they rely upon.

Archival science data traffic. Archival data must provide a well-defined quality of service (in terms of loss rates, for example), and often must meet tight standards. If the system cannot meet the requirements, it must adapt, degrading to a behavior that is still useful, for example reducing the granularity or scope of the data collection while still preserving the required QoS.

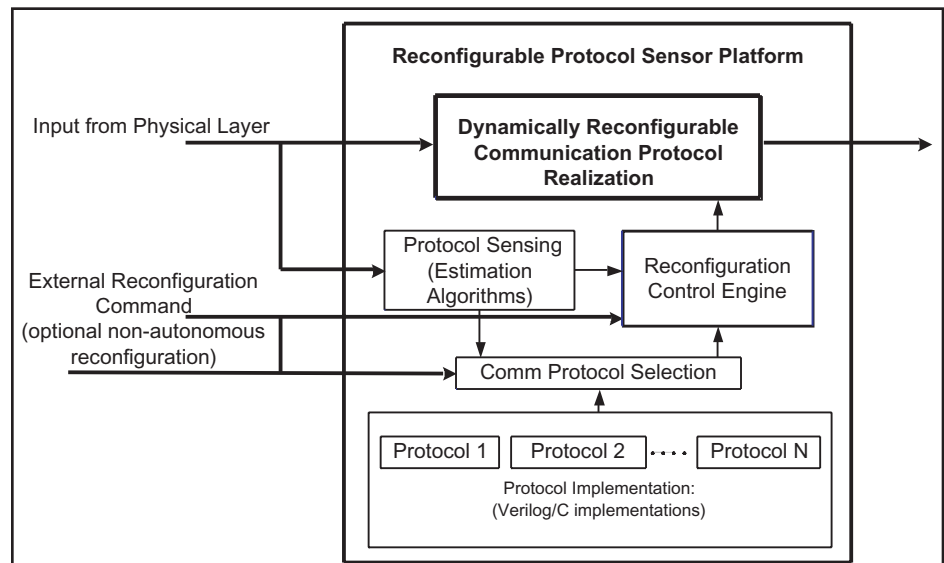
Diagnostic data traffic. In order to develop, debug, and deploy sensor systems, we anticipate a need to support diagnostic data traffic, often over the same physical media as is used by the application itself. The scale and scope of this traffic can vary, and often might be significantly scaled back after deployment is complete.

Configuration and calibration data. Calibration data, whether manually or automatically generated, can be erroneous. Examples include localization and navigation data that might have some probability of error, and sensor calibration information that might be incorrect or matched to the wrong sensor. In addition, this data will tend to become more corrupt over time: components will fail or be moved around, and sensors will go out of calibration as a result of age and environmental impact. Fielded systems will need to adapt to these cases, perhaps by detecting and rejecting data from components that appear to be miscalibrated or misconfigured.

Network configuration and topology. The sensor system must adapt to a wide variety of sensor and network configurations, including variations in topology, variations in the radio link topologies (star vs. ad-hoc), and the type and number of nodes and components.

An example of a reconfigurable communication protocol sensor platform is shown below.

Figure 4.1 Example of a reconfigurable protocol sensor platform.





Such an architecture, with its associated object-oriented design methods and partial reconfiguration techniques, enables rapid *autonomous reconfiguration* of sensor network functions in response to changes in the environment, operating conditions, and error events.

Among the challenges such a platform will overcome are:

- Enabling of error mitigation, resource sharing, and network compatibility among heterogeneous sensor networks, or networks with heterogeneous nodes.
- Enabling of reconfigurable sensor network links.
- Reducing the overall infrastructure cost of sensor networks by developing a common platform for realizing network protocols.

A similar approach can be taken to developing sensor platforms with dynamically reconfigurable sensing capability, and both platforms will immensely benefit the process of error containment and mitigation.

4.2.4 Challenges in Tools

There are many challenges in developing appropriate tools:

Automation of deployment and coverage estimation. Tools for deployment and coverage estimation will tend to be application-specific, and are intimately related to the types of sensors and types of processing needed (i.e., specific to a particular synthetic sensor). A core challenge will be taking advantage of opportunities to develop common components and tools from instances of sensor systems as more experience is gained.

Composition and configuration tools. Composition and configuration tools enable end-users to compose and configure collections of synthetic and simple sensors. A primary challenge will be to develop common interfaces and component models that will enable these tools to work across many instantiations of synthetic sensor components. These interfaces will include feedback to the user about error characteristics and confidence/uncertainty.

Validation and calibration. The challenges associated with validation and calibration extend beyond tools to the development of testbeds, data archives, and calibration procedures. While these factors will initially be highly application-specific, a major challenge will be to later decompose them into reusable tools and components. There will be limits to this process, because calibration and validation are always to some degree tied to the specifics of the analysis techniques, just as the details of

ground-truth testbeds and archived data are specific to sensors and deployment environments.

4.3 Recommendations

4.3.1 Grounding Development in Specific Existing Applications, Leading to Generalized Solutions

The ultimate goal is to develop a limited number of sensor network architectures that are tunable to a wide range of applications, taking into account that the nature of the sensor network problem requires intimate collaboration between technologists and scientists. Thus, the most viable path to achieving the ultimate goal will build on lessons learned in a carefully chosen set of specific applications. Key factors in this effort include:

- The ultimate goal: tunable synthetic sensor architectures that will apply to multiple applications.
- Method and requirements determined through dialogue between technologists and scientists.
- **Investment in ground-truth testbeds.** For sensor networks to evolve in the context of scientific applications, they must be validated in large-scale, real environments that pose all of the challenges outlined above. Validation will require reasonably controlled experiments, which implies comparing different sensor networks in the same systems, and overlapping the networks with traditional monitoring efforts. Benchmarks in the areas of sensors, networks, and coverage need to be collected at these sites and used to help critically assess future developments. Efforts of this kind will be highly resource-intensive and must be leveraged against existing and proposed large-scale field investigations.
- Derivation of benchmarks: sensing, coverage, and network.
- Developing an open source repository of reusable tools and models.
- Leveraging resources across programs.
- Validation of deployable systems.

4.3.2 Raise the Bar on Methodology and Practice

As sensor networks mature, so will the potential to make spatially and temporally rich information accessible over the Internet. Communication of this complex data must be improved to make it accessible not only to scientists, but also to policymakers, the education community, and the general public. At the same time, standards for

recording the appropriate metadata must be applied in order to preserve key supporting information. For this vision to be viable in the long run, these efforts will require a substantial, cross-disciplinary effort directed at educating and training future generations of scientists and technologists. Key components include:

- Preserving the appropriate metadata.
- Publication of data in forms that are accessible to various levels.
- Education and training of researchers in sensor and data technologies.

4.3.3 Tools: To Address the Elements Above, a Suite of Tools Will be Required

These tools fall into four categories:

- **Automation of deployment and coverage estimation.** Deployment is a difficult process, made more challenging by the growing complexity of today's systems. Tools are needed to make these systems deployable by people who are not specialists in implementation details.
- **Composition and configuration tools.** Tools are needed to enable end-users to compose collections of synthetic and simple sensors, as well as to tune the individual behavior and parameters of components, e.g., latency requirements. This process of composition must expose information about error propagation.
- **Validation and calibration.** Tools are needed to enable validation and calibration of a sensor system. Some aspects may be performed in the lab while others must be performed in a prototype deployment, perhaps including comparison to ground-truth measurements. The analysis techniques embedded in synthetic sensors may need to be modeled, simulated, and validated before application in new environments.
- **Metadata tools and requirements.** There is a need to preserve metadata describing the sensor characteristics, error characteristics, and uncertainty associated with archived data.

4.4 References

- [1] F. Koushanfar, M. Potkonjak, and A. Sangiovanni-Vincentelli. Fault tolerance in wireless ad-hoc sensor networks. *IEEE Sensors*. 2: 1491-1496, 2002.
- [2] S. Slijepcevic, S. Megerian, and M. Potkonjak. Location errors in wireless embedded sensor networks: sources, models, and effects on applications. *ACM Mobile Computing and Communications Review*. 6(3): 67-78, 2002.
- [3] W. Munro, C. Thomas, I. Simpson, J. Shaw and J. Dodgson. Deterioration of pH response due to biofilm formation on the glass. *Sensors and Actuators B-Chemical*. 37(3), 187-194, 1996.
- [4] S. Marrs, R. Head, M. Cowling, T. Hodgkiess and J. Davenort. Spectrophotometric evaluation of micro-algal fouling on marine optical windows. *Estuarine, Coastal, and Shelf Science*. 48(1), 137-141, 1999.
- [5] S. Wainright, J. Kremer and C. D'Alvanzo. Evaluation of Endeco 1184C dissolved oxygen recorders for use in temperate estuaries. *Water Research*. 29(9), 2035-2042, 1995.
- [6] A. Sciortino, T.C. Harmon and W.W-G. Yeh. Experimental design and model parameter estimation for locating a dissolving DNAPL pool in groundwater. *Water Resources Research*. 38(5), U290-U298, 2002.
- [7] A. Sciortino, T.C. Harmon and W.W-G. Yeh. Inverse modeling for locating dense nonaqueous pools in groundwater under steady flow conditions. *Water Resources Research*. 36(7), 1723-1736, 2000.



Chapter 5. Security

5.1 Introduction

While distributed sensor networks have great potential for advancing science, distributed collections of environmental data carry significant security implications. Sensor network architects and users must address security issues from the initial system design, and continue to do so with the data collected well after the network is dismantled. In a general sense, most security problems found in distributed sensor networks are also found in other distributed computer systems. However, the embedded nature and scale of distributed sensor networks pose novel security threats and exacerbate others.

Examples from the Internet motivate the need for investment in privacy and security. Consider the large amount of data generated and posted publicly on the Internet in the 1990s, without concern for security or privacy. At the time, lack of explicit control was of limited risk because data were transient, difficult to search, and seen by relatively few people. However, the data were archived, and are now indexed and easily searchable by today's search engines. Similarly, in the 1980s and early 1990s, systems attached to the Internet were rife with security vulnerabilities, but exploitation of these holes was rare and piecemeal. Today, in contrast, even a single vulnerability can cause widespread economic disruption.

Analogues to these and other problems exist in sensor networks. Data collected from a sensor network today may be difficult to exploit and seemingly innocuous. However, future improvements in programmability and data mining may result in unintended consequences. It is also clear that sensor networks can be attacked, which will result in erroneous data being saved. Future networks comprised of millions of embedded sensors might even provide a platform for a network or physical attack.

Users of sensor networks have security needs that are similar to users of traditional systems. They need data integrity and authentication: they want to know that the data they receive are uncorrupted, and know where they came from and when. Networks must maintain availability and be resilient to disruption; sensor networks that do not produce data are not useful. Privacy is needed, both for the scientists and the objects being observed. For reasons of correct attribution of work, scientists must be able to perform experiments confidentially, prohibiting others from viewing experiments in progress. There is also an issue of privacy regarding certain data that may inadvertently contain information beyond what the

experimenters sought to gather. And while these needs fit into well-understood security categories, their threats and the means to neutralize those threats do not.

Key sensor network vulnerabilities include denial of service attacks, passive listening, and data insertion or corruption. Denial of service [1] can occur in many ways (e.g., by physically inserting a device that jams the wireless communications). Since a distributed sensor network may be deployed in remote regions, an adversary may physically destroy some subset of the devices. The wireless communication also permits passive listening by unauthorized individuals. Even worse, the insertion of corrupt sensor or control data could cause the system to stop operating, operate dangerously, make the collected data meaningless, or cause incorrect data to retard or wrongly direct scientific investigation.

Data collection on a large scale can have unintended consequences that can cause security risks. For example, a large system deployed in the ocean, such as NEPTUNE (<http://www.neptune.washington.edu/>), can use microphones and sonar to monitor fish migrations. However, these raw data may unintentionally record faint traces of the U.S. submarine fleet; an adversary may be able to mine the raw data to learn valuable military intelligence.

The issue of data mining also poses threats to people's privacy. For example, once many sensor networks exist, data from different systems might be merged and assessed to acquire unexpected information about individuals, corporations, or governments. People need some degree of understanding and control over how they are observed by such networks, allowing them to make informed decisions about their privacy.

5.2 Challenges and Solutions

Three key factors pose significant security issues and challenges distinct from those found in traditional Internet-based systems: scale, embedment, and privacy [3]. As scientists and researchers deploy greater numbers of large-scale sensor networks, the security requirements of these systems and their impact on these three factors will become clearer. Identifying and characterizing these new security models is a significant task.

Sensor networks exist at many scales, from the 50-node NEPTUNE network to mote-based networks with thousands of nodes. Even larger systems and systems-of-systems will exist in the future. This wide range of scale

imposes a correspondingly wide range of security challenges and required solutions.

5.2.1 Challenges

Modern computing systems such as laptops and desktops are typically rich in computational resources: they use billions of CPU cycles and hundreds of megabytes of memory to edit text or view images. This growth in power has allowed what were once computationally taxing operations to become commonplace. For example, when Adelman, Shamir and Rivest first proposed RSA encryption in 1978, encryption with a cutting-edge VAX computer took on the order of 30 seconds. Today, RSA encryption is used every time a secure website is accessed, taking a few milliseconds. These techniques may be applied to wired, resource rich nodes such as NEPTUNE.

In contrast, mote-based sensor networks are resource limited. With processors only marginally faster than those of a 1978 VAX and a few kilobytes of memory, they cannot afford to use the same algorithms and mechanisms that have become commonplace on personal computers. Since 1978, however, the importance of security in computing systems has increased greatly. For example, the first Internet worm was ten years later, in 1988. Mote-based sensor networks must meet modern security needs but have available only limited resources, e.g., current motes must solve security problems with resource capacities similar to those available in general purpose processors twenty years ago.

In addition, mote-based networks are composed of large numbers of devices. A mote network administrator may be responsible for thousands of devices, and keeping track of each individual node is not feasible. As the scale of the network increases, this decreases the mean-time-to-failure of a node from the network. In networks with a large number of nodes that can readily fail, the administrator focuses on maintaining operation of the network as a whole even with these problems. The security model of a mote-based network must be similarly resilient to failure. This broad range of scales for networks results in a spectrum of security approaches, and heterogeneous networks must deal with many points on that spectrum simultaneously.

Unlike traditional computing systems, sensor networks are embedded in uncontrolled environments. For example, in Internet-based systems such as Web servers, physical compromise is rarely an issue, as the computers are in dedicated and locked server rooms. In sensor networks, however, the opposite conditions generally prevail, and nodes are not similarly protected. Instead, the network is often deployed in remote locations, far from easy visual

observation. Under such conditions, an adversary can physically compromise nodes even if the network communication is secure, and systems must be able to continue to operate in the presence of compromised nodes.

Not only does embedment pose security risks to a sensor network, it also raises questions on security implications for the collected data. Monitoring the environment can lead to gathering data on unsuspecting (or unwilling) subjects. For example, as mentioned above, the U.S. military has recently been concerned with NEPTUNE's deployment of seismographic and acoustic sensors in the deep ocean. Although the sensors are intended for geological, chemical, and biological research, the same data could be used to monitor ship and submarine movements. Protection against unintended uses of data is a very challenging problem.

As a result of the special needs of sensor networks, new security models must be developed. New metrics for assessing the security and safety of these systems are required. Fundamental questions that relate the lower bound on resource requirements necessary to meet various types and degrees of security need to be answered. Means to assess the impact of compromised nodes on the final accuracy of the collected data must be developed.

5.2.2 Solutions

The following proposed solutions are not meant to be exhaustive, but rather to illustrate directions that can provide some immediate solutions.

Many nodes used in sensor networks provide limited resources for computation and communication. These limitations severely hinder the use of widely available implementations of cryptographic algorithms that have driven security solutions in the broader community [5]. Research aimed at developing *light-weight implementations of cryptographic algorithms* [2] could enable for sensor networks a large collection of techniques that have been tested and evaluated in a broader community.

Given a sensor network consisting of thousands of nodes operating in a harsh environment, node failures due to factors such as hardware errors, software bugs, or attack are inevitable. In addition to securing individual nodes, it is necessary to design systems that are *resilient to attacks* and other forms of node failure. The concept of graceful degradation has been a cornerstone of distributed and fault tolerant systems, and the applicability of this approach to sensor networks and security should be explored. In particular, systems should be able to continue to operate in the presence of compromised nodes. The broader community has developed a number



of approaches for detecting intrusions and network anomalies. These approaches may be fruitfully adapted to the environment presented by a sensor network. Such approaches should make it possible to *identify compromised nodes* and *revoke any rights* they may have within the network. As an example, work in wireless ad hoc networks that enables each node to actively overhear the wireless channel, identifying anomalies of its neighbors' transmissions, has demonstrated the capability of such active defense to be an effective counter to attacks [6,7].

Physical compromise of a sensor node could reveal critical information (e.g., encryption keys) that could be used to impersonate the compromised node. Special, *tamper resistant* nodes that destroy their storage upon physical tampering would defeat such an attack.

Characteristics of the deployed network and the subjects being sensed can be used to validate the authenticity of collected data. As an example, identifying the presence of an automobile in one location at one instance followed immediately by an indication that the automobile had moved a great distance or that the automobile was following a physically impossible path could be an indication that the network is being spoofed. Also, given the high density of sensors in networks, the inherent redundancy can be exploited to solve some of these security problems.

The correct operation of middleware services such as the localization of nodes, time synchronization, data routing [2], and self-calibration are essential to the functioning of many sensor networks. When necessary, these *middleware services* should be secured against attack. A number of proposals [2,10,11] have begun to address these issues, but the broader space of such problems remains largely unexplored.

Attacks can be launched against different levels of a system. A malicious "black-hole" node might try to attract data from nodes throughout the network, interfering with the data-collecting ability of a real base station. A "jammer" might transmit noise to disable the communication in its vicinity. *Multiple layers of defense* not only protect the network from a diverse spectrum of attacks, but also ensure that a breach of one line of defense does not compromise the entire system.

Sensor network users are likely to perceive security as an absolute, i.e., they are likely to believe that the system is either secure or not secure. As with other systems, the reality is not so well-defined. A sensor network may be protected from some security violations while being vulnerable to others. Specific issues include the degree of trust and the potential for social impact (e.g., invasion of privacy) of the sensing and data collection activities.

Scientists and the public need to be informed about the complex consequences associated with deployment of sensor networks. This aspect of security is best addressed through education.

In practice, sensor networks are likely to be deployed by scientists who are not security experts. A *composable security infrastructure* which supports the construction of sensor networks from smaller parts that are secure and trusted will be invaluable to the future deployment of sensor networks. As an example that works for the Internet, SSL (Secure Sockets Layer) provides an infrastructure that allows individual machines to be added to the Internet while retaining the desired security properties.

Future sensor networks may require large numbers of heterogeneous nodes. Authentication schemes will need to be able to scale to the magnitude required to support such large-scale systems. The building blocks of *authentication* should have sufficient *modularity* to easily enable interoperability among heterogeneous software and hardware components for a coherent system.

5.3 Recommendations

We recommend that funding agencies support or initiate activities in three general areas: (1) basic research in cyber security, influenced by the unique characteristics of sensor networks; (2) the development of prototype or testbed sensor network systems that have security as an essential component; and (3) educational activities ranging from the education of scientists on issues of ethics and security to public outreach on the role and impact of sensor networks.

5.3.1 Basic Research in Cyber Security

While it is clear that the security challenges introduced by sensor networks will benefit from general research in cyber security, sensor networks present four research opportunities that are unlikely to arise in other contexts. First, the security of sensor networks should take advantage of properties of the physical environment in which they are deployed. This exploitation of physical properties to enhance network security is a fertile ground for novel techniques and mechanisms. Second, security mechanisms of sensor networks should self-organize to minimize human intervention. Because of the potentially large scale of sensor networks, autonomous approaches such as self-diagnosis and self-healing are necessary to relieve the user from the burden of attending large numbers of nodes individually. Third, research should identify the extent to which not just individual nodes but overall system architectures can be secured.

Because many sensor networks will be constructed from sensors with severely limited resources, traditional approaches that emphasize the security of individual nodes may not be appropriate. System level approaches, including resilience techniques that ensure operation of the network in the presence of a certain percentage of compromised nodes, should be investigated. Finally, because sensor networks rely on the correct operation of specific services such as routing, localization, etc., research should investigate the degree to which the security of these “middleware services” can be enhanced, in light of the limited resources available on a sensor node.

5.3.2 Testbed Sensor Network Systems

While many of the issues related to security in sensor networks can be studied in isolation, design and implementation will need to be examined in a more complete context. To ensure the validity of approaches to network security, funds are needed to support the development

of fairly large testbed/prototype sensor network systems that involve multidisciplinary teams from both science and technology. These systems should be driven by scientific exploration of a specific phenomenon where security is an explicit requirement. Security must also be an integrated part of the design from the beginning.

5.3.3 Education

In the traditional education sense, we need to educate the next generation of computer scientists and engineers who will design and implement security solutions for sensor networks. Scientists need to be cognizant of the fact that deployed sensor networks may be capable of unintended observations, and the consequent privacy implications. This may involve developing specific guidelines regarding the deployment of sensor networks. Outreach programs should be designed to help the public become informed of the policy-related issues.

(Resources continued on page 38)

Box 8. Fixed Ocean Observatories



Figure 1. Cable Node Highlights: NEPTUNE’s 3,000-km network of fiber-optic/power cables will encircle and cross the Juan de Fuca tectonic plate. Between 30 and 50 experimental sites will be established at nodes along the cable and will be instrumented to interact with physical, chemical, and biological phenomena that operate across multiple scales of space and time. Node locations shown here are hypothetical. Final decisions on placement will depend on science input and engineering considerations.

In order to study ocean related phenomena, a fixed cabled backbone provides both power and a data path for long term experiments and the support of autonomous vehicles (Figure 1). Power levels to the experimental nodes can be as high as 5 kW and data rates of 5 Gb are planned. These systems provide instrumentation nodes for the connection of multiple sensors at multiple geographic locations involving a wide range of scientific disciplines (Figure 2). The details of a system must be established during the design phase with the understanding that changes over the 30-year service life are extremely expensive and probably not practical (Figure 3). In addition, the total system life cycle cost is driven by not only the initial fabrication and installation cost, but the number of repairs required over its life. Repair requires the mobilization of a vessel capable of recovering the system element, its replacement and re-installation. These operations are expensive and expose the system to additional damage. Once the system has been installed, there is essentially no way to make change or fix problems.

The problems that must be addressed during the initial design include:

- A data and power backbone compatible with a wide range of sensors
- Mechanical configuration suitable for deployment and repair from an affordable vessel
- Adequate reliability to meet the available life cycle cost
- A repair strategy compatible with the obsolescence of hardware over 30 years

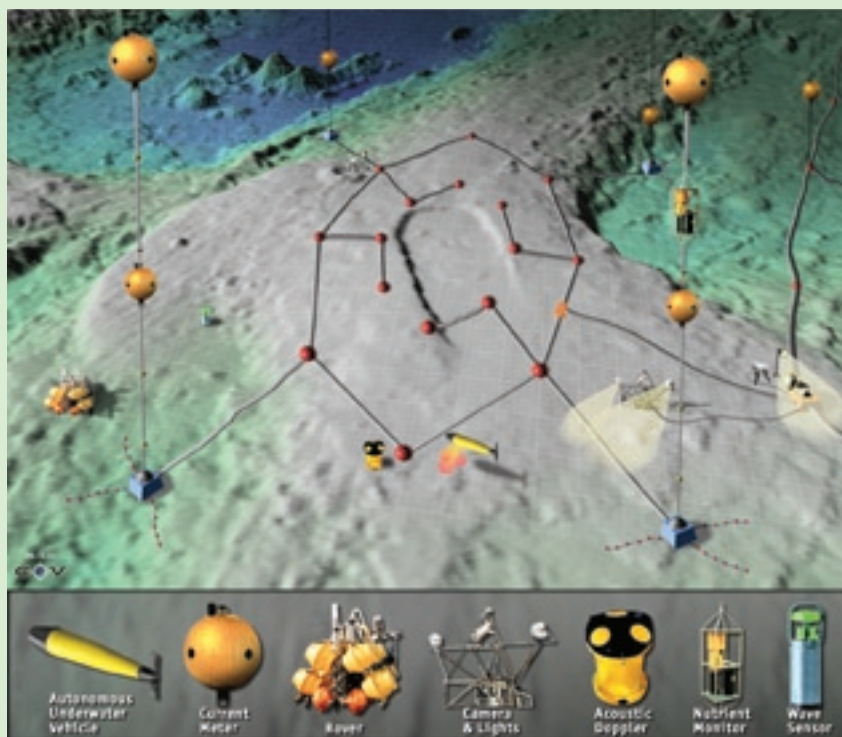


Figure 2. Essential Elements: Land-based scientists, educators, decision makers, and the general public will be linked via the Internet to sensors and sensor networks in the water column, on the seafloor, and in the subseafloor. The NEPTUNE infrastructure is being built to have an expected lifetime of 30 years and will serve as a community resource, much like a research vessel is an observational platform open to a range of users.

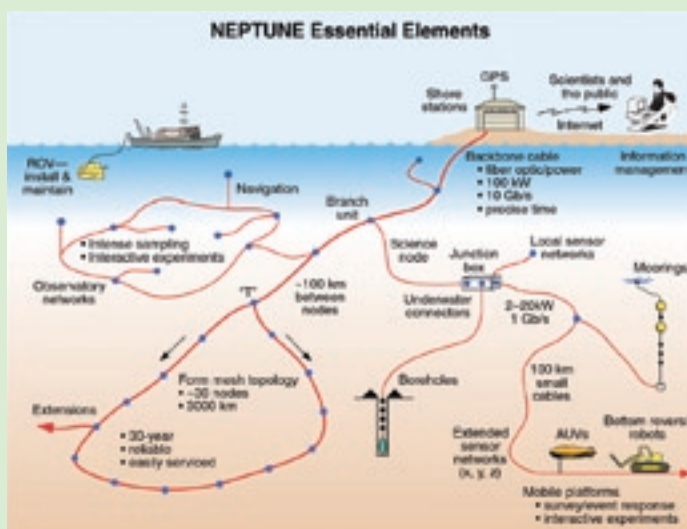


Figure 3. Axial Seamount: Shown here is a generic NEPTUNE experimental network draped over Axial Volcano and based on the National Oceanic and Atmospheric Administration/Pacific Marine Environmental Laboratory's New Millennium Observatory (NeMO). The network will provide real-time command-and-control capabilities to shore-based users via the Internet. Autonomous underwater vehicles will reside at depth, recharge at nodes, and respond to events such as submarine volcanic eruptions. This image is representative of the kinds of installations that might ultimately be located at each of the experimental sites.

The data and power backbone should be designed to eliminate or control the potential interference between the switched electrical components and the sensor inputs with all of the elements imbedded in the same ocean. The arrangement of the cable, the nodes and the science instruments must include considerations of cable deployment and recovery. The system reliability includes initial topology, redundancy, fail-soft modes, part selection and pre-installation test. Long-term repair of the electronics must accept the changes in technology over a long period of time.

The keys to the successful completion of these programs include:

- The establishment of a program manager with adequate staff at the onset of the program
- A fully funded system engineering group to use models and tradeoff studies to translate the science requirements into affordable hardware specifications
- Installed prototype system to uncover design problems
- A competent detailed design, fabrication and installation contractor with experience in and an understanding of ocean systems

Images provided courtesy of the NEPTUNE Project (www.neptune.washington.edu) and CEV (<http://www.cev.washington.edu/>).

5.4 References

- [1] A. Wood and J. Stankovic. Denial of service in sensor networks. *IEEE Computer*. 15(4), 48-56, 2002.
- [2] V. Wen, A. Perrig and R. Szewczyk. SPINS: Security protocols for sensor networks. *Proceedings of the seventh annual international conference on mobile computing and networking*. Rome, Italy, July 16-21, 2001. pp 189-199.
- [3] L. Zhou and Z. Hass. Securing ad hoc networks. *IEEE Network*. 13(6), 24-30, 1999.
- [4] D. Pescovitz. A big radio in a (very) small package. *Lab notes*. 3(3), 2003. <http://www.coe.berkeley.edu/labnotes/0403/spec.html>
- [5] D. Carman, P. Kruus and B. Matt. Constraints and approaches for distributed sensor network security. *NAI Labs: Technical Report # 00-010*, 2000.
- [6] S. Marti, T. Giuli, K. Lai and M. Baker. Mitigating routing misbehavior in mobile ad hoc networks. *Proceedings of the sixth annual international conference on mobile computing and networking*. Boston, MA, August 6-11, 2000. pp 255-265.
- [7] H. Yang, X. Meng and S. Lu. Self-organized network layer security in mobile ad hoc networks. *Proceedings of the first ACM Workshop on Wireless Security (WiSe)*. Atlanta, GA, September 28, 2002. pp 11-20.
- [8] S. Basagni, K. Herrin, E. Rosti and D. Bruschi. Secure pebblenets. *Proceedings of the second ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc)*. Long Beach, CA, October 4-5, 2001. pp 156-163.
- [9] C. Karlof and D. Wagner. Secure routing in sensor networks: attacks and countermeasures. *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications*. Anchorage, AK, May 11, 2003. pp 113-127.
- [10] Y. Hu, A. Perrig and D. Johnson. Ariadne: a secure on-demand routing protocol for ad hoc networks. *Proceedings of the eighth annual international conference on mobile computing and networking (Mobicom)*. Atlanta, GA, September 23-26, 2002.
- [11] J. Deng, R. Han and S. Mishra. A performance evaluation of intrusion-tolerant routing in wireless sensor networks, *Proceedings of the second international workshop of information processing in sensor networks*. Palo Alto, CA, April 22-23, 2003. pp 349-363.
- [12] A. Perrig. The BiBa one-time signature and broadcast authentication protocol. *ACM Conference on Computer and Communications Security*. Philadelphia, PA, November 6-8, 2001. pp 28-37.



Chapter 6. Data Management

6.1 Introduction

Environmental data are intrinsic to our ability to understand the workings and assess health of the biosphere [1,2,3]. New technologies and advances in sensor technologies make it imperative to have an appropriate data management model to address the emerging challenges and opportunities as new data streams come on-line from sensors deployed within a framework spanning multiple temporal and spatial domains. It is essential to adopt a fundamental systems approach to all aspects of data and its management, because without reliable, abundant multidiscipline data it will not be possible to discover the new knowledge about the biosphere, how it works, and the transitions it is undergoing—knowledge that will be vital both to science and society.

Data are frequently lost or mis-managed when handled in ad hoc ways, or are stored but remain unused because without appropriate data infrastructure, a data warehouse can become a data graveyard. Changing media technologies also represent a significant problem as valuable data stored on older technologies (paper tape, magnetic tape, Hollerith cards, Bernoulli drives, 8, 5.5, 3.5 inch disks, and so on) must be systematically migrated before they are lost.

In addition to the technological issues, there is an important human dimension to data management. Significant cultural barriers exist to adopting modern practices, and these barriers must be overcome in order to make sensor networks effective tools. We describe these barriers and suggest that a combination of educational initiatives, new technologies, and policy changes by funding agencies and scientific societies will accelerate the success of sensor networks and the management of resulting data. We also believe that this paradigm shift is ripe for study by social scientists interested in processes of change within scientific disciplines.

6.2 Challenges and Solutions

Our increased ability to collect data has led to the “swamped in data” problem, involving both volume and heterogeneity as well as legacy data. As new sensor systems are put in place, concerns about being overwhelmed by data and related management challenges is becoming an urgent problem. The Internet and other data collection technologies have opened up vast resources, including data sets of all types and increasingly larger sizes. There appears to be no end to

the production of new data sets as technologies advance and organizations recognize that data can be generated and reused in multiple ways. A systems framework is required for designing an appropriate data management model that will support the production of the knowledge required for both deeper scientific understanding and reliable guidance on environmental policy. As we enter the new realm of data collected from the environment using both fixed and mobile sensor arrays, we must address the question: **How do we best manage and archive the growing high-bandwidth heterogeneous physical, chemical, and biological data streaming from sensing systems?**

Volume of data. As high speed wireless networks are put in place to provide communication from the field to the laboratories, new sensor networks will emerge with increased data flow and accumulation. Key technologies developed within the sensor network itself may help the volume issue. For example, data aggregation solutions will not only help to address issues of volume but will also help solve practical problems of power management in the sensor network. It is unlikely that it will be feasible to have for a continuous stream of all the data pour out of the sensor network into a backend database.

Presently, field stations and marine laboratories archive only a fraction of the data that they collect or could collect. Remoteness of the field sites, low-bandwidth communications to universities and archive centers, poor data management infrastructure support, and culture barriers all contribute to the problem. Consequently, many stations and researchers sample only a tiny fraction of the data that the sensors are capable of collecting. The new generations of sensor arrays and the resulting data they will produce will only exacerbate the problem unless new data models are adopted.

Heterogeneity of data types. The measurement of environmental data in the field ranges from use of multi-spectral scanners aboard spacecraft and measurements of temperature to the acquisition of streaming sound and video. Data types include camera images and video, infra-red measurements of organism activity, protein analysis for organism identification, chemical measurements (e.g., CO_2 , NO_x , SO_2 , CH_4), pheromones, sonar, microwave, radar, acoustic signals, storm events, and water flow, to name just a few. Multi-spectral satellite imagery at multiple scales of temporal and spatial resolution provides new information to assess the status of ecosystems.

(Continued on page 42)

Box 9. NEON—National Ecological Observatory Network

The vision laid out by the scientific community for the National Ecological Observatory Network (NEON) is bold and ambitious [1]. A nationwide network of field and laboratory-based research facilities spanning gradients of environmental factors and human influence, NEON will provide the unprecedented ability to tackle mounting environmental challenges that increasingly demand an understanding of biological and ecological phenomena over large spatial and temporal scales. By coordinating research and data collection across the observatories in the network, NEON will foster new approaches to ecosystem research that encompass the inherent complexity of the environment. Each observatory's regional footprint will be defined by the arrangement of an intensively studied core site and several satellites where specialized field-based sensors will monitor a host of biological, chemical, and physical processes that control the structure, composition, and dynamics of ecosystems in the region (Figure 1). Collaborations among diverse research and educational institutions will provide laboratory space as well as the human capital needed to operate the observatories and coordinate research among them.

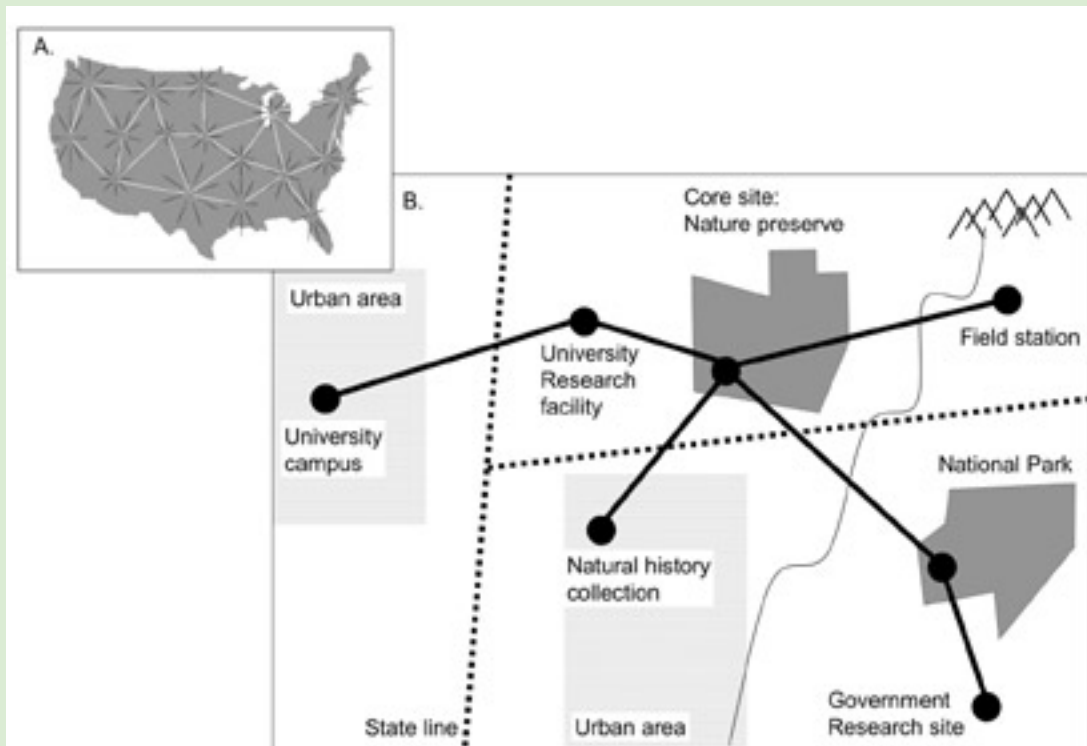


Figure 1. A. National and B. Regional NEON footprint



At the heart of NEON lies cyberinfrastructure—the assemblage of sensors, databases, analysis applications, data portals, and other technological resources that will permit the collection, processing, integration, analysis, modeling, and dissemination of high-quality ecological information (Figure 2). NEON will support and coordinate integrated measurements in the following areas: climate and hydrology; biodiversity and population assessment; biogeochemistry; biosphere-atmosphere coupling; and spatial analysis and remote sensing. Thus, NEON will not only need to leverage existing cyberinfrastructure, but it will also require the development of both sensors that provide novel and efficient ways to explore ecosystems and new schemes for converting the massive data streams they produce into useful environmental information. NEON is made possible by the recent information technology revolution. However, the ultimate realization of the vision relies, in large part, on continued advances in the development of environmental cyberinfrastructure.

References

[1] K. Holsinger and the IBRCS Working Group. IBRCS White Paper: Rationale, Blueprint, and Expectations for the National Ecological Observatory Network. Washington, DC: American Institute of Biological Sciences, 2003.



Figure 2. NEON Cyberinfrastructure

The diversity and complexity of the data from sensor networks will challenge data management specialists.

Legacy of datasets and legacy datasets. Important ecological phenomena and processes (e.g., human impacts on ecosystems, carbon sequestration, greening of the land and waters, changing biodiversity, etc.) occur across large temporal and spatial scales. The emergence of landscape and global change ecology over the last few decades is testament to the recognition by ecologists that these large scales deserve attention. These long time and large spatial scales suggest that data collected by sensor networks will have value for many years to come. Incorporation of historical data sets with emerging sensor network data will be essential for testing many specific hypotheses and for developing forecasting models.

Data archiving strategies: domain specific or collection specific. A natural difficulty in designing a data management system is how such information can be archived, managed, and made accessible. For instance, if an ecologist collects micrometeorological data along with environmental acoustic data, it makes sense that this information should be stored together or linked via relational database management technologies as a part of a collection using a storage structure that will allow efficient packing and access. Alternatively, one might argue that micrometeorological data could best be curated and discovered if it were stored in some facility managed by the atmospheric science community, provided it is made accessible using interoperable technologies.

Environmental sciences: a collection of sciences practicing without metadata or standards. Environment science encompasses a wide variety of disciplines, traditions, scales, and instrumentation. At one end of the scale there are field biologists working alone and needing little more than a water-proof notebook and a good writing implement, while at the other end we find atmospheric scientists working in large teams with automated networked sensors and sophisticated computational models. In the tradition of individual science there is little need for metadata or disciplinary standards because scientists communicate by publishing papers, while in the large team disciplines people are more specialized in their roles for the collection, transforming, analysis, and reporting about data as well as maintaining the data collection and management systems. Sensor network science requires that individual scientists be willing to relinquish having complete control over the process in order that measurements can be made at larger and longer scales. Central to this process is the sharing of data, and this can only be enabled with meta-

data. At present, environmental science still lacks a core metadata standard.

Data management understanding by students of environmental science. Environmental science students have little experience in the management of large datasets. Although data are the driving component in discovery science, knowledge of proper management of large, diverse datasets is generally lacking. To understand how to manage and access the valuable data provided by arrays of diverse sensors that is urgently needed by the environmental science community, it will be imperative for students and researchers in environmental science to gain a systems perspective that encompasses all aspects of data, from collection *in situ* to the end product of knowledge generation. A system science approach is imperative because only in this way will it be possible to manage all aspects of the data life cycle. As ecology makes the transition to larger and larger geographic scales, the need for new educational paradigms to educate the next generation of environmental scientists is becoming critical.

Ready access to data is limited. Ready access of data depends on our ability to find, select, download, and use the appropriate data on the Web. Each of these steps currently presents problems for end users. Many of us now rely on search engines such as Google to find digital information, but because data sets are most efficiently stored in databases, and databases cannot be readily indexed by search engines, a great many valuable data sets often remain difficult to discover.

The goal of distributed software applications is to use registration servers to overcome the discovery problem. A simple application such as USA Photomap can obtain black and white JPEG images of the US land surface over the Web in a fairly intuitive way from the TerraServer using a friendly user interface, but generalizing this process to multiple sources is a problem that is only now being addressed in the Web services community, and its solution depends on a number of building blocks that are not yet in place (e.g., UDDI for discovery, APIs for querying databases, high-speed connections to download data, and user friendly applications to view and analyze data). Engineers in the business and scientific communities are only now developing prototype systems to address these problems.

Lack of coherent data models. There is a tension in science between the desire to collect more data in the hope that new phenomenon will emerge (discovery mode) and hypothesis-driven experiments and subsequent data collection [4, 5]. As with other promising new instruments, we anticipate that environmental sensor



networks will allow scientists to discover new phenomena. It is critical to develop specific data models that are motivated by specific objectives. Everyone needs to know why the data are being collected. Initially there may be a greater focus on understanding the building blocks of sensor network systems design, but as the building blocks become better understood we will be able to iterate model development and efficiently tune the control system to respond to the hypotheses being tested. Such controls as adjusting sensor sampling rates, sensor sensitivities, and network topology, and communication bandwidth are sure to play important roles. System design will require the separation of physical and logical representations of the network and may require a phenomenological language.

6.3 Recommendations

- **Design data model flexibility in hardware and software systems.** One of the significant issues associated with the design data model associated with sensor networks is data fidelity. Due to power, storage, and bandwidth constraints, data reduction, data compression, and data archiving issues are at the forefront of concern. Yet many environmental scientists want and need access to raw data. Thus, the concept and practice of data modification at the sensor is troublesome. Design of a data model for sensor networks must include data systems that do not place limits on hypothesis testing or knowledge generation due to technological constraints.
- **Development of standards to facilitate data exchange and multidisciplinary research.** With several exceptions in the environmental sciences, history, experience, and the reward system tend to focus personal effort on small, discipline-specific communities in which the individuals review each other's work and compete with one another for funding. Thus, we carry around a neighborhood view of the world. The need to coordinate common large resources in large data networks and aerial image resources in Atmospheric Sciences and ships in Oceanography has necessitated building a scientific community that can share data. It is likely that sensor networks will have the same effect in the environmental sciences.
- **Integrating the development of large scale information based on prototype systems.** To make the environmental sensor networks successful, scientists need to marry the sensory arrays to communications and storage networks. Examples that exist or are being built include HPWren and RoadNet. These examples can provide useful guidance as environmental scientists begin to expand their sensor networks.
- **Interdisciplinary prototyping and testing (end-to-end).** One of the grand challenges in systems of automated networks for the collection of multiple variables from the environment is the need to understand the processes from end-to-end. This cannot be accomplished by single investigators or by uni-discipline teams. Sensors are complicated and sensor technologies are complex. To be successful in accomplishing environmental sensing, environmental scientists must understand the full dimensionality of the data stream from measurement variable, sensor placement, sensor calibration, data reduction/compression, data flow, data archiving, data management, data access, visualization and analysis, and interpretation, to delivery.
- **Broad usability of data systems outside the development community.** Sensor networks are a community resource that will take decades to design and build. They will play a central role in informing public, managers, and policy makers about the state of the environment. Like other large scale infrastructure facilities, it is essential for the developer community to have a clear vision about how this infrastructure will provide the answers sought by the wider community, and for the scientific and wider communities to have realistic expectations about the outcomes of such efforts. Forecasting and providing warnings about biological events are compelling mandates, but the complexity of biological systems suggests that it will require a concerted effort to make such forecasts dependable.
- **Educating the next generation environmental scientists about data use from sensor networks.** Universities are rapidly developing new approaches to environmental science and policy, and are enhancing existing programs to address emerging global issues. They are adopting new technologies to enable faculty and students to address environmental issues. But there is a general lack of understanding about how to use emerging technologies such as complex sensor networks and the data that will be produced by multiple arrays of environmental sensors. There is no data model that can be adopted to address the myriad types of data that will become accessible to the environmental science community.
- **Take a life-cycle view of data and data management.** Data management in almost all cases we have encountered is characterized by an ad hoc approach, individual decisions, poor archive practices, and passive data loss [6]. In contrast to published papers where there are library systems for indexing and archiving synthesized data, the information embedded in data itself has historically not been valued. The

challenge now is to take a “life-cycle” view of data, in which scientists and the library community apply a systematic approach to how data can best be managed.

- **Interact with other scientific and industry communities (e.g., NASA, Physics community, Oceanography, Open-GIS, ISO TC 211) to capture the best practices.** Many organizations are in the process of addressing the increase in data flow and volume and the associated management challenges their communities face. Public and private organization has invested billions of dollars to streamline information flow and management. We particularly want to develop strong collaborative linkages with standards organizations (e.g., Open-GIS). Wherever possible, it may be useful to adopt open source approaches, which provide reusable building blocks and promote interoperability.
- **Education and outreach related to data management.** One of the driving needs for gathering multi-dimensional data using scaleable sensor and communication technology is to provide relevant information to planners, decision makers, and politicians. Collecting larger amounts of data at higher temporal and spatial resolution will be more fully supported by the public when the data leads to knowledge that helps solve such looming environmental problems as terrestrial, aquatic, and atmosphere pollution; land use change; the destruction of ecosystem services; sustaining the quality of life; and maintaining biodiversity. The cyberinfrastructure community must strive to develop information systems that can be accessed and utilized by decision makers, enabling them to make science-based decisions that are supported by knowledge derived from data and information. Effective education and outreach efforts can inform and engage the public with the importance of the services provided by healthy ecosystems. As technologies advance, new paradigms can be developed to enhance environmental awareness. One example is to make available in people’s homes the sounds of the environment, enabling the public to select different sites at different seasons and times of the day, and to hear the heartbeat of ecosystems and become aware of ecosystem changes and health.

6.4 References

- [1] J. Lubchenco, A. Olson, L. Brubaker, S. Carpenter, M. Holland, S. Hubbell, S. Levin, J. Macmahon, P. Matson, J. Melillo, H. Mooney, C. Peterson, H. Pulliam, L. Real, P. Regal, and P. Risser. The Sustainable Biosphere Initiative: An ecological research agenda. *Ecology* 72(2): 317-412, 1991. <http://esa.sdsc.edu/91annualrep.htm>
- [2] J. Lubchenco. Entering the century of the environment. A new social contract for science. *Science* 279:491-496, 1998.
- [3] H. Schellnhuber. ‘Earth System’ analysis and the second Copernican revolution. *Nature* 402: supp c19-c23, 1999.
- [4] J. Meredith. Theory building through conceptual methods. *International Journal of Operations and Production Management*. 13(5), 3-11, 1993.
- [5] D. Huron. Lecture 3. Methodology: The new empiricism: systematic musicology in a postmodern age. *The 1999 Ernest Bloch Lectures*. University of California, Berkeley. Department of Music, 1999.
- [6] W. Michener, J. Brunt, J. Helly, T. Kirchner, and S. Stafford. Non-geospatial metadata for the ecological sciences. *Ecological Applications*. 7: 330-342, 1997.



Chapter 7. Metadata

7.1. Introduction

Metadata are information about data. Library catalog systems are a familiar example of metadata about books (e.g., author, title, publication date, etc.). Currently, in environmental science there is a large and growing amount of data, accompanied by relatively scarce information describing and classifying that data. Growth in electronic and information technology and its digital products has accelerated interest in accessing, retrieving, and exploring data, creating demand for ways of better understanding the data that are available. Archiving comprehensive metadata together with data is one important approach to solving this problem.

In the context of sensors and sensor networks, *metadata* are formally structured documentation related to the data collected by sensor networks, for example, the type of sensor, its location, dates, sampling rate, etc. A *schema* is a formal definition of the structure of a data set. In traditional computer databases, a database schema defines the entities in the database and the relationships among them. More specifically, it spells out the types of data items (e.g., a “name” field and an “address field” both hold strings), and the logical structure of the database (e.g., a “customer” consists of a name and address). A *metadata schema* is a formal description of the types of metadata stored to describe a data set, and the relationships among them.

Metadata add overhead to data management efforts. Further, metadata creation, if done manually on a sensor-by-sensor basis, can become a major task when large networks of sensors are deployed.

The combined scientific audience that builds and uses environmental sensor networks is a diverse and multidisciplinary group. There are many parties, with distinct interests and knowledge bases, working on disparate problems using similar sensors and technologies. What may be accepted practice in one field may be foreign to another. This multidisciplinary nature makes the tasks of metadata creation and use highly complex.

A major problem with making metadata useful to the broadest possible community is that there is no single agreed-upon method in the scientific community for generating metadata. There are, however, various standards that can be used and applied in sensor deployments to facilitate metadata creation. On the other hand, there are relatively few standards pertaining to scientific data content, especially across multiple disciplines.

While the concept of metadata is straightforward and the tools available are useful (if not fully adequate), few people within the research community are formally trained in metadata creation and use. While training workshops exist (e.g., the U.S. Geological Survey’s National Biological Information Infrastructure programs) a serious deficiency remains in educating data publishers about how to produce and use metadata in conjunction with existing digital resources.

A related issue is the lack of incentives for generating and archiving metadata. In general, data are collected by a specific group with a specific end-use in mind. Within this group, the collection methods and purpose are clear and known, and thus metadata may be perceived as having little internal value. But if the data are archived and/or later transferred to another group, the original background context in which the data were gathered may be lost. Since generating metadata is not without cost, there is little incentive for the originating group to spend resources on publishing metadata since they will receive little benefit from it. Further, at present little recognition comes from being a good citizen and voluntarily creating metadata that is useful to and benefits the wider community. Both of these issues raise barriers to the effective creation and use of metadata for subsequent analysis and wider use.

7.2. Challenges and Solutions

7.2.1 Technical Obstacles to Successful Metadata Collection

There are several different classes of problems that need to be addressed in order to improve the quality and use of metadata. In the vision where thousands to millions of environmental sensors, deployed by multiple disciplines for specific research goals, are providing data and data products, the description of the data and the decisions as to what information to store for query and archival purposes will be decisive in determining the long-term value of these data.

From a data-gatherer perspective, one fundamental challenge is the need to develop metadata input systems that can scale by orders of magnitude along with increasing numbers of sensors. In keeping with this is the need for instrumentation to be self-describing, so that the metadata can be updated as new instruments are connected to the sensor network. A second challenge

is to develop metadata systems that can evolve as new innovations are implemented in making measurements. A third challenge is to develop descriptions for sensor network state-of-health, data quality assurance, and quality control. It is clear that there are (and will continue to be) rapidly evolving standards and requirements that will demand the accurate tracking of sensor and network changes in close to real-time.

From a data-user perspective a significant challenge is to identify and access all of the appropriate metadata necessary for using sensor network data. There is a need to develop methodologies that will allow extensions and additions to metadata as new requirements develop or new data products come into existence. One of the more difficult challenges is posed by legacy data. How can the metadata and data from legacy systems be brought forward into the emerging world of fully accessible state-of-the-art data, overcoming the potential energy barrier that is impeding this important step?

7.2.2 Disciplinary Obstacles to Successful Metadata Collection

The collection of metadata in a form useful to as wide a user constituency as possible is crucial. The scientist consumers of sensed data derive great benefit from being able to identify and use data sets that are related, and may come from multiple disciplines. For example, a scientist performing analysis and/or modeling using one data set may seek other data sets gathered from sensors placed in a similar environment in order to compare the data themselves and to compare the results of an analysis or cross-validate the predictive power of a model. It is the metadata for a data set that allow an investigator to identify how data sets are related and whether they are suitable for use together. Moreover, once related data sets are identified, accurate and detailed metadata are essential to integrating distinct data sets—data collected using different instruments and under different operational conditions will require carefully designed processing steps to render them directly comparable.

Unfortunately, in practice there is no shared framework for describing metadata across different scientific disciplines. Absent such a framework, interchange and integration of data sets gathered by investigators in different disciplines (e.g., soundshed, oceanshed, watershed, airshed, etc.) are either difficult or impossible. Even within a discipline, many communities have yet to define standard schemas for the recording of metadata.

7.2.3 Social Obstacles to Successful Metadata Collection

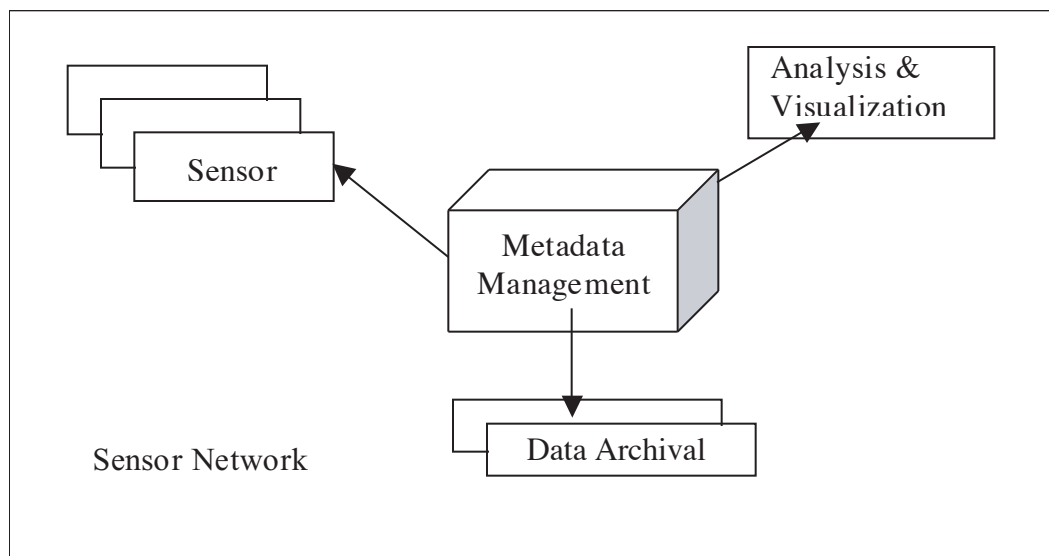
There is a divide in the scientific community between investigators who prefer an open-data policy, where data are made available to the entire research community, and a closed-data policy, where each investigator privately retains data. This divide can be characterized as social: those who do not share data most frequently choose not to because of concerns about receiving publication credit for the research results derived from the data. Advocates of open data, on the other hand, often view widespread distribution, use, and scrutiny of data sets as an efficient mechanism for improving confidence in a data set's validity. A wider community of data consumers is more likely to catch anomalies in data, and thus identify problems with a fielded sensor system. Those who believe in the closed-data model often pay little attention to gathering detailed metadata *in a form that will be useful to others*, because they have yet to adopt the model of sharing data with others.

Implementing the practice of annotating data sets with attribution metadata is a key challenge. Investigators may be more inclined to share their data if the *data and any published results based on them* carry a clear source attribution. If data were digitally watermarked, they would be indelibly branded with their contributor's identity. NASA encourages those who use NASA data to cite their source. This policy not only encourages sharing of data by giving credit to those who publish it, it also encourages careful gathering and archiving of metadata in a widely usable form, to maximize the use of a data set by other research groups.

Another social phenomenon is the gap between data providers and data users. Often, the engineers who design instruments and deploy fielded sensors have a different view of metadata than the scientists who consume the data. For example, instrument designers are frequently concerned with the details of the instruments themselves: calibration and algorithms for post-processing raw data to correct for instrument limitations. Scientist consumers of data, however, have different concerns (e.g., ambient conditions not captured in the data themselves) at the time a set of data was recorded. The limited degree of interaction between instrument designers, authors of tools for managing data and metadata, and users of data and metadata perpetuates this perspective gap. The schema for metadata must take into account not only the instruments' physical characteristics and the deployment's characteristics, but also capturing information relevant to how the consumers of the data will use the data. More interaction between these communities must occur to achieve this goal.



Figure 7.1. Metadata management as a core component of a sensor network.



7.2.4 Solutions—Establishing Standards

There are a number of potential solutions to the challenges raised in this chapter. Key to these solutions is to realize that metadata management is a core component of any sensor network, and that the sensor network (like any sensor) is embedded in a workflow for scientific investigation, for which metadata management is equally crucial. This realization is summarized in Figure 7.1 with respect to the sensor network and in Figure 7.2 with respect to the scientific workflow.

The nature of sensor networks is such that we must look at the overall scientific process and the role of metadata in that process. Effective support for metadata management within sensor networks demands that we have more effective support for metadata management in the broader scientific community. In turn, this requires a fundamental review of how scientific information is collected and managed. We need to move from an environment where much of the fundamental metadata (e.g., measured quantity definitions, units of measure definitions, models of natural phenomena) is kept in books (electronic or otherwise) to an environment where this metadata information is accessible via online machine-readable registries that can be integrated into the practice of the working scientist.

In order to meet the needs illustrated in Figures 7.1 and 7.2 (next page) standards will be needed. Such standards are required for an overall framework for metadata management as well as discipline-specific standards for metadata content.

Let us first consider the standards for an overall metadata framework. What are the requirements for such metadata frameworks?

- **They must be metadata schema-independent.** The framework must allow usage of any metadata schema in order to deal with the diversity of measured quantities, processing services, phenomena, and sensor types that are required. Metadata frameworks that are only suited to the description and discovery of data sets are unacceptable. Metadata must be attachable to any component of the scientific process; sensors, measured quantities, data transformations, phenomena, and statements of scientific problems.
- **The framework must be extensible.** Objects in the sensor network are not static, and the metadata system must be able to deal with “late in the game” realizations regarding what are important metadata parameters.
- **Metadata schema development must be extensible in a hierarchical fashion.** We should all be able to share common definitions for things like floating point numbers, strings, and dates. Most of us will also want common representations for geographic objects, location, extent, coverages, coordinate reference systems, and time. We will want to build on the basic types in creating content models in specific domains. In addition, it must be the case that the things that we create (i.e., the vocabulary of our own domain) be shareable and serve as input to the development of vocabularies in other domains.
- **The framework must provide for life cycle management of the metadata.** Metadata resources are going to come and go and change in terms of their priority.
- **The framework must be Web accessible.** It must be possible to both search and update the metadata within a wide area network (e.g., the Internet).

Figure 7.2. Metadata management in the scientific workflow

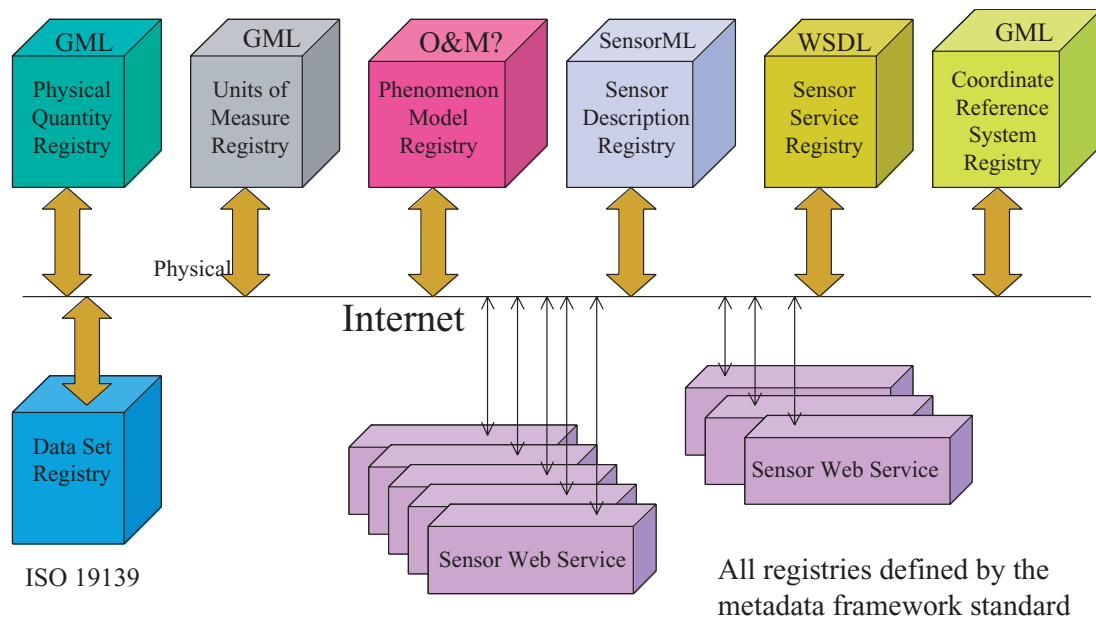
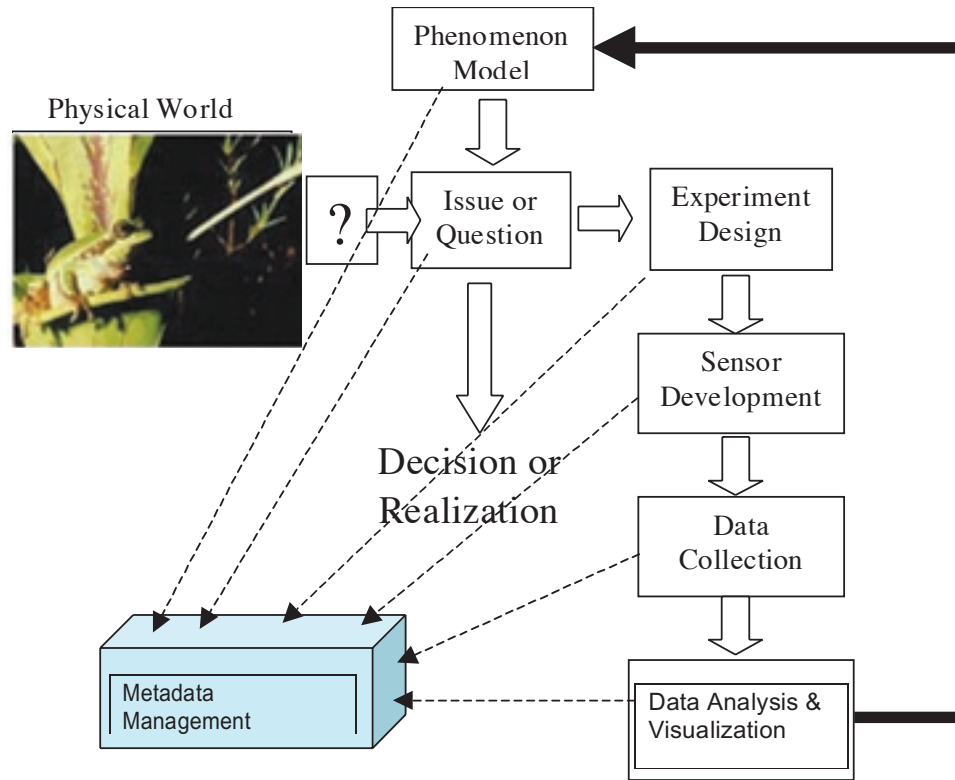


Figure 7.3. Registries for sensor networks. The Physical Quantity Registry provides formal, machine-readable definitions of measured quantities (e.g., Sea Surface Temperature). These definitions, in turn, depend on (are associated to) units of measure definitions (which may contain conversion formulas, etc.), and are used by phenomenon models maintained by Units of Measure and Phenomenon Model registries respectively. All of these registries are shared resources and would be used by a practicing scientist in the description and analysis of scientific experiments, in experimental planning, and in the design and development of sensing devices. It is anticipated that sensor manufacturers will make use of these registries to submit sensor descriptions to the Sensor Registry.



- **The framework must be a Web service in the general sense.** This means that the Web savvy software must be able to query and update metadata over the Internet. This will enable the community to develop tools to integrate metadata capture into the scientific workflow and thus greatly reduce current impediments to metadata acquisition.
- **The framework must provide for internal integrity checks and audit trails.**
- **The framework must support fine-grained access control.** Some users will be able to update specific metadata records, others will be able to change metadata schemas, while still others will only be able to browse and read metadata records. Such sophisticated access control is essential for data integrity, user confidence, and practical sharing of metadata resources.
- **The framework should provide the ability to link or associate metadata resources with one another in an unrestricted manner.** Data processing algorithms may be associated with particular kinds of measurements, sensor descriptions may be associated with sensor data services, data sets may be associated with data access Web services, and so forth.

Development of metadata content standards is also crucial. Such content standards, must, however, be developed using a common metadata framework. This is essential to sharing the metadata content and for resolution of overlaps and conflicts between different content standards. Only where we have a common, shareable metadata representation will we be able to resolve the critical semantic issues relating to the integration of metadata content over a range of scientific disciplines.

A concrete implementation of the metadata framework is in the form of a registry/repository, and standards have been developed in other domains (e.g., OASIS, eBRIM) that can be applied to the scientific process and sensor networks. We believe that specific types of such registries, dependent on specific types of metadata schemas, need to be deployed. These registries serve as the foundation for more specific metadata registries to be developed for vertical application domains. Figure 7.3 (left) shows a set of these fundamental registries. The letters at the top of each box provide a suggested metadata framework language that could be used to develop the metadata schemas for that registry.

7.3 Recommendations

Many past and present attempts to develop and incorporate metadata standards have failed to achieve full success due to the resistance of data providers to support the effort, the lack of “buy-in” by the complete range of

end-to-end participants, and the development of numerous incompatible standards by narrowly-focused communities. The following section addresses some possible solutions for improving success of data and metadata standards within an environmental sensor network.

7.3.1 Decreasing Resistance/Increasing “Buy-in”

- Providing incentives for data collectors to make their data not only accessible but usable by a large community of potential users (citing data use in publications, recognition by NSF and peer reviewers of the importance of data providers, sufficient funding to support proper data management and adherence to metadata standards).
- Providing tools that decrease the inertia of supporting metadata implementations (wizards to assist in sensor description, metadata and data encoding, software libraries that assist developers in incorporating data and metadata parsers within their software).
- Provide standards that the community can trust (e.g., sufficient backing, authoritative overseer, built upon industry standards).
- Engaging the entire end-to-end chain of players from the very beginning of the process and throughout testing and early adoption.
- Providing standards that affect a large enough community to make them profitable for commercial entities to support.
- Push for “buy-in,” not “force-in” by identifying measures that make it beneficial to data suppliers, users, and tool developer to be a part of the effort (e.g., financial support, demonstrably better results, easier workflow).

7.3.2 Engaging End-to-End Community

Ultimately, the successful adoption and use of any metadata standards require “buy-in” by all players within the end-to-end data chain. This includes the sensor providers, data collectors, data managers, Web service providers, visualization and analysis software developers, and ultimately the data user and decision maker. If any of these players is not fully engaged in supporting the standard, then the potential success of these standards can be greatly diminished.

In order to fully engage all players in the data chain, it is important to be aware of the potential incentives of each, as well as the responsibilities and challenges of each, as shown in Table 7.1 (next page).

Table 7.1. Achieving metadata “buy-in”. Possible incentives and responsibilities of communities within the end-to-end data chain needed for successful adoption and use of metadata standards

Community	Incentives	Responsibilities
Sensor Providers	Larger sales market; More sensors sold	Need to provide initial sensor descriptions within the standard sensor description framework and adhere to standard sensor interfaces
Data Collectors	Data will be more widely used; More recognition for work; More funding received	Commitment to supplying complete high-quality metadata in accepted standard; Need to obtain support from sensor supplier or provide support themselves
Data Managers	More recognition/funding/profit; More time-efficient and cost effective operation	Need to support the metadata within search and query capabilities and provide “hooks” for Web service providers
Web Service Providers	Greater use of services; Increased market and profit for services, if commercial; More recognition and funding, if non-profit	Need to develop Web service engines and Web services based on these standards
Visualization and Analysis Software Developers	Ease of development, more functionality, ease of use; Larger potential user community; Larger market/profits if commercial, more recognition/funding if non-profit	Need to provide end-user tools that not only are able to parse standard data and metadata documents but are also capable of utilizing this information to provide improved functionality
End Users and Decision Makers	Better results in a more timely manner and with greater ease; Ability to easily investigate new multi-discipline questions and relationships not previously possible or practical	Need to educate themselves about the potential of these data and metadata; Need to demand adherence to and use of these standards by tool suppliers

A key factor in assuring end-to-end buy-in for these metadata standards is to engage each group as early as possible in the process, starting with the design, development, implementation, testing, and adoption stages. If this is not done, there is significant risk that metadata standards designed by data suppliers or end users will fail to fully utilize existing or emerging industry standards, which will significantly impede their adoption by key service and tool providers in the middle of the data chain.

Engaging end-to-end players in the process from the beginning also greatly increases the sense of ownership by all and the desire for all parties to assure its success. In addition, allowing all parties to begin to implement and test these standards early in the development process provides several benefits: (1) unforeseen problems are exposed early where they can be more efficiently corrected; (2) confidence in the standard is increased; and (3) sensors, data, tools, and services that utilize these standards are available before or immediately after adoption of the metadata standards.

7.3.3 Improving Inter-community Use of Metadata Standards

Too many different “standards” tend to be developed within the scope of a narrow community. This results in a large number of standards that are incompatible and not well supported by tool and service providers.

To understand the problem, one must recognize the difference between content standards, framework standards, and encoding standards. A well-designed content standard is based on fundamental concepts about the physical data or sensor output, and may, for example, specify that the description of a particular observation should contain at a minimum properties such as the physical phenomenon represented, geospatial position, time, units of measure, and data quality. However, the content standard does not specify how this information should be organized or encoded. These data models, if well designed, often remain relatively stable, even if framework models or encoding of this information changes. The content standards are best defined by the sensor or science communities who have the most expertise regarding these data.



A well-designed framework model may also remain fairly stable, even though the means of encoding this model may change. A framework standard provides a means of organizing content under a common philosophy and structure. A successful framework design provides a robust, general, extensible model that can effectively support a variety of content. Currently, these models can be specified using the Universal Model Language (UML), which provides a visual means of portraying information.

Finally, these content and framework standards must be encoded to provide a means to write, store, transport, and access the actual data. Traditional means of encoding have included Excel spreadsheets, binary files that required either documentation or software libraries to understand the content, or ASCII files with key-value pairs. Currently, more web-appropriate encodings include the World Wide Web Consortium (W3C) eXtensible Markup Language (XML) and Resource Description Framework (RDF).

Unfortunately, metadata standards are often designed so that the content models are too closely coupled to one framework or encoding. This has the potential to decrease the acceptance and longevity of the standard, since the entire effort may be lost if underlying framework or encoding is no longer viable or compatible with evolving industry standards. This has been the fate of numerous standards that were no longer suited for the web-based paradigm.

Utilizing standard metadata frameworks for dynamic geospatial data provides these benefits:

- Science community efforts can be focused on maximizing the expertise of the community. That is, the scientific community can focus on providing standard semantics and data content that are relevant to that community, rather than trying to develop an entire metadata framework that will be useful only to a small community.
- Science communities that adhere to a higher-level metadata framework will be able to more easily utilize data from other science communities, as well as data from other non-science data sources that adhere to these standards.
- Sensor manufacturers, data managers, Web service providers, and end-user software developers will be more apt to support these standards since it broadens their market to a larger number of communities without requiring additional development.

In summary, to improve both the utility, interoperability, and longevity of standards developed for environmental sensor networks, it is recommended that the science and sensor communities focus on defining the data content models, and that these content models be incorporated into frameworks that have been developed to support a larger scope (e.g., those defined by ISO TC211, the OpenGIS Consortium, or other geospatial standards groups). The encodings of these frameworks should in turn be based on international industry standards defined by bodies such as W3C, ISO, and IEEE.

Box 10. North Temperate Lakes Monitoring

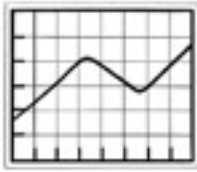
Many ecological systems are characterized by high spatial and temporal variability [1], non-linear dynamics [2], and coupled physical/biological processes [3]. This combination of traits often results in complex spatial and temporal patterns of ecological processes and phenomena. Understanding the causes and consequences of these patterns presents major scientific challenges that are both ecological and technological in nature. The difficulty in collecting, managing, and analyzing data sampled at varying frequencies (minutes to weeks) at locations distributed widely across a landscape hampers our ability to observe, let alone understand, complex spatial and temporal dynamics. Recent advances in smart, networked arrays of field-deployed sensors offer new promise for collecting ecological data at these scales. However, building, operating, and optimizing these systems require the collaborative efforts of experts in multiple disciplines. Integrating the hardware capable of sensing and communicating key ecological data in a power-limited environment, and designing an information management system to interpret signals and control the operation of the sensor arrays are areas of active, cutting-edge research.

At the North Temperate Lakes Long-Term Ecological Research site in northern Wisconsin, researchers have deployed a series of instrumented buoys on remote lakes to capture data at frequencies as high as one minute (Figure 1). Data are automatically transferred over bi-directional wireless transceivers every hour, loaded into an Oracle database, and made available in near real-time on the web (<http://lter.limnology.wisc.edu>). Current limitations include power-hungry sensors and communications infrastructure, range in license-free radios, sensor calibration, and a deficit of intelligent command systems that adaptively turn on and off sensors or change the frequency of sampling depending on environmental conditions. It is critical that the solutions to these challenges scale to embedded networks within lakes, among lakes, and across lake districts. Smart, networked, instrumented buoys offer great promise to uncover high-frequency data across extended spatial scales, a time and space regime currently very poorly understood.



References

- [1] T. Kratz, L. Deegan, M. Harmon, and W. Lauenroth. Ecological variability in space and time: insights gained from the US LTER program. *BioScience*, 53:57-67, 2003.
- [2] S. Carpenter. *Regime Shifts in Lake Ecosystems: Pattern and Variation*. Volume 15 in the Excellence in Ecology Series, Ecology Institute, Oldendorf/Luhe, Germany, 2003.
- [3] D. Hamilton and S. Schladow. Prediction of water quality in lakes and reservoirs: Part I - Model description. *Ecological Modelling*, 96: 91-110, 1997.



Chapter 8. Analysis and Visualization

8.1 Introduction

Tremendous growth in both sensor network technology and anticipated applications is driving the need for new techniques and tools for analyzing and visualizing sensor network data streams. These data streams have the following characteristics:

- They are massive in dimensionality, spatio-temporal extent, and aggregate rate.
- They are highly heterogeneous and multi-modal.
- They consist of, or must be integrated with, data with widely varying temporal dynamics and spatial attributes.
- Increasingly, they will be derived from networks that include mobile sensing capabilities.

These traits involve new degrees of complexity in analysis and visualization methods. Offline methods, including spatio-temporal pattern recognition, event detection, decision support, and knowledge discovery, must grapple with all of these characteristics, and integrate seamlessly with the systems for data and metadata management.

We can also anticipate important applications in real-time analysis and visualization in support of control, adaptation, and resource management. Each of these must operate in support of the realistic characteristics of the sensor network itself: processing constraints, variable communication latencies, and limited energy budgets.

Multiple types of sensors or multiple sensor arrays may be used to observe the same physical phenomenon, but may not have been designed to work together or may use vastly different data representation mechanisms. Scientists will need to be able to correlate observations from among a number of diverse sensor arrays.

Finally, it is clear that sensor networks will have multiple users. While scientific inquiry often drives first-generation applications, distributed sensor networks have the potential to arm both policymakers and the general public with powerful information that vastly enriches science-based policymaking, education, and public awareness. Lack of awareness, limited accessibility, and poor ease of use continue to be major barriers to data use by non-specialist communities. For these reasons, it is imperative that tools be designed with multiple user communities in mind.

These characteristics pose new research and development challenges for a wide spectrum of researchers from almost every science and engineering discipline. A number of anticipated and developing application domains are already shedding light on the new challenges ahead:

Ecogrid. Ecogrid is an effort to construct a national grid-based infrastructure that applies the framework and services of grid computing for long-term ecological research in Taiwan. It includes five sites encompassing five different kinds of ecosystems across the island, ranging from subtropical mixed evergreen hardwood forest to seashores and coral reefs. Ecogrid integrates a hierarchical wireless network infrastructure with data acquisition, sensor control, and robotics (<http://ecogrid.nchc.org.tw/>). The national research network of Taiwan will be used as its backbone, in which high performance computing, storage, and visualization resources are connected and served in a grid manner. The goal is to leverage cyberinfrastructure to enable a large number of integrated ecological research projects. This project serves as a model for future systems that employ massive, multi-modal, heterogeneous data streams and a variety of user communities.

Flows of biota. While our capabilities to measure physical phenomena such as atmospheric microclimate and marine chemical species are improving greatly, the ability to sense and quantify ecosystem processes currently limits our ability to understand system-level phenomena. There is a critical need to improve the measurement, visualization, and analysis of the timing, magnitude, and rates of transport of organisms by multiple transport processes. Such flows of organisms have implications for the prediction of diseases; the distribution and spread of exotic or invasive species; changes in migratory pathways and food resources; changes in abundance of flows of organisms; and continental-scale flows of genes. The need to quantify and understand dynamic ecosystem networks, as opposed to environmental phenomena, poses a broad set of challenges. A critical subset of those challenges includes those related to the research communities involved in data analysis and visualization.

Smart farming. The goals of smart farming are to optimize productivity and efficiency, monitor the health of plants and animals, and minimize pollution. The sensed data sets are used for real-time analysis and actuation and off-line analysis. Thus, analysis tools for this application must span scales from minutes or hours to years.

In addition, the sensed data may have multiple users beyond the individual farmer, including nowcasting (short-term meteorological prediction) and hydrology.

Health monitoring. Sensing the health conditions of humans and animals can help evaluate the effects of pollution, improve data collection in epidemiological studies, and improve response times for emergency medicine. The collected data streams are diverse and multi-modal. Examples include real-time heartbeat characteristics and long-term tracking of metabolic and biochemical parameters. Finally, security and privacy are both paramount concerns in this application.

8.2 Challenges and Solutions

8.2.1 Theoretical

In order to realize the full potential of distributed, heterogeneous sensor networks that may potentially be deployed on a global scale, significant advances are required in the state-of-the-art of both theory and algorithms for distributed estimation, detection, and decision-making in bandwidth- and energy-constrained environments. This need is complicated by a gap in understanding the interrelationships between underlying phenomena occurring at diverse temporal and spatial scales. These, in turn, induce complex dependencies across sensors that defy strong modeling assumptions. Classical methodologies for data analysis and visualization are insufficient to this task. Therefore, a fundamental challenge is to develop both theory and algorithms for robustly aggregating and visualizing information from a network of heterogeneous sensors across space, time, and sensing mode. It is unlikely that methods which do not embrace the full complexity of the problem will be able to achieve this advance.

Some of the challenging characteristics of sensor network data include dynamic statistical dependency structures, latencies between measurements, communication constraints between sensors and processors, and finite energy resources. Furthermore, time-critical applications utilizing real-time analysis and visualization must account for network-induced control and communication latencies, competition for sensor network assets, and data exfiltration in a shared, resource-constrained environment.

It is well known that even in the case where the statistical model, implied by the complex and dynamic “graph” structure associated with a sensor network topology, is fully specified, exact inference is intractable. Additionally,

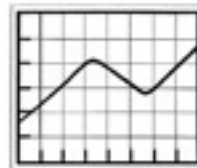
it is generally the case that statistical models across heterogeneous sensors are lacking or that the model for a given sensor network is only partially specified. The former introduces issues of the scalability, computability, and tractability of approximate inference, while the latter highlights the need for principled machine learning and adaptation methods for developing models from sensed data. The combination of these factors highlights the need for integrated statistical approaches for distributed analysis and visualization methods.

Non-invasive, high-bandwidth sensing technologies require the manipulation and processing of large-scale, high-resolution data and data sequences from multiple modalities (e.g., visual, acoustic, seismic, infrared, etc.). Visualization is complicated by both the massive quantity and diverse properties of the data. The ubiquitous aspect of sensor network applications highlights the need for visualization capabilities to scale across many different devices. Users must be able to comprehend and integrate multiple, high-rate information and knowledge streams. This is particularly important in applications such as emergency response and decision- and policy-making. Finally, sensor networks will increasingly exhibit complex and dynamic physical topologies, requiring multi-scale geo-referenced visualization. Collaborative efforts to develop analysis and visualization tools are currently limited by incompatibilities in existing data and metadata formats.

8.2.2 Development/Engineering Challenges

Analysis and visualization can play a major role not only in the use of distributed sensor networks in scientific discovery, resource management, and public awareness, but in the design of sensor networks themselves. A significant challenge is to develop engineering practice for the design of complex sensor network architectures emphasizing the interdisciplinary applications of such systems. Since modern sensor networks will be hierarchical, large, temporally dynamic, and heterogeneous, analysis and visualization tools can enable and strengthen design and optimization approaches. These approaches will include the consideration of issues such as changing models, new phenomena, component longevity, and energy consumption, and robustness/redundancy tradeoffs. Given the potential that tasks may change significantly over the life of a sensor network, developing principled engineering practice for adapting and modifying existing sensor network infrastructures poses an additional challenge.

There are also clear gaps in the translation of theoretical advances into widely available, well-documented tools



and toolsets for analysis and visualization. These gaps include:

- **Tool complexity.** Many tools are currently designed by specialists, and are thus difficult to use, modify, and maintain. Moreover, effective toolsets are becoming too difficult to build from scratch.
- **Lack of awareness and availability.** Due to limitations in current publication and discovery tools, it is difficult for non-specialists to discover and obtain relevant tools and data.
- **Incompatibility.** Tools are often developed on an ad hoc basis without the benefit of compatibility with the evolving national cyberinfrastructure of grid computing.

8.2.3 Strategic Partnerships Challenges

The end-to-end objective of using sensor nets to record phenomena that are then subsequently correlated, analyzed, and interpreted in scientific investigation is an enormous undertaking. The entire pipeline represents a large and complex system that must be designed and engineered through collaborations among numerous specialized disciplines. Most scientific and policy questions are not limited to single community or single nation impact. Biological and ecological systems, for example, simply do not stop at borders. Border can be interpreted as national, discipline, or community. A central challenge is to recognize and support interactions across all of these diverse borders.

For example, security systems in large casinos use literally thousands of video surveillance cameras to observe and reconstruct events. Observations of ecological reserves (e.g., HPWREN/Santa Margarita) use a small number of video cameras. In both cases, automated recognition of “events” from video is needed, and what are seemingly unique requirements in one discipline can cross the border and become shared needs. There are tremendous resources available in other communities (tools, systems, data, and human expertise). A key challenge is to understand how to share capabilities technically among communities, where standards should be defined, and how to enable such communication.

An example forum in which some of these types of interactions can occur is the Pacific Rim Application and Grid Middleware Assembly (PRAGMA). It is an open organization established to build sustained collaborations and to advance the use of grid technologies in applications among a community of investigators working with leading institutions around the Pacific Rim.

8.2.4 Education, Outreach, and Training Challenges

Significant education, outreach, and training are necessary at all levels of society and across scientific disciplines in order for both policy makers and the general public to exploit the full benefits of sensor networks. Within the scientific community there is the question of how to train the next generation of scientists across many disciplines in the use of complex analyses and visualizations. The complexity of existing tools and methods as well as those developed in the coming years will inevitably come with a steep learning curve, requiring an interdisciplinary approach to such training. A necessary step is to develop both the technologies and policies for putting data in the hands of the broad scientific community. Tools and methods also need to be improved for presenting complex data sets to those outside the specialized scientific community, including policymakers, emergency crisis managers, and the general public. Compelling and timely visuals that convey complex information in an informative and succinct manner will be key. For example, some local and state governments provide forecasts and current air quality data. NOAA (National Oceanic and Atmospheric Administration) eventually plans to issue forecasts of ozone and particulate matter for the whole nation. Wise use of graphics and standardized terms to notify the public will be needed, much as is now the practice with weather warnings. Clear and timely dissemination of data can be immensely improved in many other disciplines (e.g., stream levels, mosquito populations, traffic congestion, and drought conditions). One of the most urgent needs is the rapid and effective collection and presentation of data in civil emergencies such as toxic releases, terrorist attacks, and earthquakes. Development of data algorithms and visualization tools for clear presentation of data related to such events is important to a large portion of the population, and can greatly increase the benefits realized from the data that are collected by sensor networks and made available.

8.3 Recommendations

To encourage and enable advances in analysis and visualization research we propose recommendations in four areas: (1) theory—algorithms and data structures; (2) development/engineering—building tools and systems, and establishing standards and policies; (3) strategic partnerships—forming multidisciplinary and international collaborations; and (4) education, outreach, and training—informing policy makers, students, and the public.

8.3.1 Theoretical Research Recommendations

- **Identify challenge problems** to frame and motivate sensor network research (e.g., quantifying and analyzing flows of biota at multiple scales). Integrate these problems with similar programs for air pollution.
- **Support algorithm development** in statistics, machine learning, and visualization for complex sensor network applications, emphasizing the distributed aspect of phenomenology, analysis, and data collection.
- **Create analysis and visualization tools** that employ image and signal processing, incorporate spatial contexts, and enable synchronization. These tools should enable processing and interpretation of high-bandwidth sensor streams.
- **Include visualization capabilities** using a range of devices, from high-end technologies such as tiled display walls and CAVE interfaces to small displays on personal digital assistants and mobile phones.
- **Develop multi-scale geo-referenced visualization**, including integration with GIS techniques and remote sensing imagery.
- **Develop new display systems** that integrate high-resolution imagery and video, high-fidelity audio, and tactile interfaces to support virtual and augmented reality environments. The visualization of complex phenomena could also benefit from alternate sensory representations, such as the use of audio cues for the interpretation of complex temporal or spectral data sets.
- **Develop a common family of visualization data formats** to ease data fusion, support tool integration, and encourage collaboration between users and user communities. Ad hoc metadata approaches could be adopted in earlier development phases, with future development consolidating these approaches with domain-specific customization in a modular manner.

8.3.2 Development/Engineering Recommendations

- **Develop analysis and visualization tools** for new methodologies that will enable the design of future sensor networks. Efforts should also be promoted that advance simulation-based design of hierarchical sensor systems that are constrained in energy consumption, bandwidth, and environmental/ecological impact. New systems should integrate sensor network

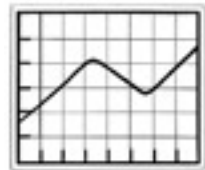
architectures and visualization/analysis tools. In addition, new approaches should be sought that enable distributed, real-time control and planning of network function and resource allocation.

- **Develop scalable frameworks and techniques** for the development of analysis and visualization toolsets that target new functionality in knowledge discovery and dissemination. Framework development efforts should be integrated with educational curricula and programs that develop a new breed of researchers and practitioners with expertise in both scientific application domains and the advanced analysis and visualization tools.
- **Develop a services registry and underlying services architecture** (Web and grid services) for data and tools to support publication, discovery, and access. Resources should be allocated to provide training and education in the design of tool components, their assembly into toolsets, and their use by a broad range of users.
- **Support the development of experimental test-beds and prototype systems.** Where appropriate, these systems should capture common interdisciplinary problems and the strength of international collaborations.
- **Establish and evolve open frameworks** for the development and dissemination of interoperable toolsets, including the means to adapt tools from other problem areas.

8.3.3 Strategic Partnerships Recommendations

- **Support (international) workshop series** to enable collaborations and communications among different applications and technology groups. In addition, fund efforts to make sustainable the collaborative opportunities that are identified.
- **Encourage and fund formal standardization and community standardization efforts** in data representation, data mark-up, and metadata standards.
- **Examine and fund efforts to create solutions that are applicable across several domains**, and identify areas where domain-specific extensions need to be made.
- **Create strategic cross-disciplinary partnerships** with complementary communities, for example, among biosurveillance, law enforcement, and homeland security; between financial and telecommunications industries; and between medicine and bioinformatics.

(Continued on page 58)



Box 11. Observing the Acoustic Landscape

Stuart Gage and his students at Michigan State University have developed the “Clickable Ecosystem” with a focus on recording acoustic signals in different places. A small computer system is developed to automate recording acoustic signals, weather from a wireless weather station and images from a web camera and automatically transmit information to a remote server. Acoustic signals are recorded at half-hourly intervals each day and sent to a data server [1] via wireless, broadband, DSL or satellite communications. In some places where networks are not available, data are downloaded from disk weekly and transferred to the file server.

The signals are compartmentalized to compute the intensity in the frequency domains in the signal. Biological and landscape indices are computed based on ratios of acoustic intensity in each of eleven 1 KHz frequency bands. The acoustic signals, derivatives from the analysis of the signals and ancillary observations associated with the acoustic signals (temperature, precipitation) and images are placed into a digital library using relational database technology to provide access to the acoustic signals the analysis of them. A web tool [2] provides remote access to the digital library where all of the acoustic signals and ancillary data are stored. Forty-eight acoustic signals are automatically recorded each day from each place, thus providing an in-depth acoustic signature from locations monitored. These acoustic signals and the synthesis of the frequency elements can be accessed by the Web to enable the public to select different places at different times of the day to hear the heartbeat of ecosystems and allow them to assess ecosystem health. The digital library of acoustic signals provides a rich accessible database to examine sounds of humans, other organisms and the physical actions in the environment.

References

[1] Computational and Ecology Visualization Laboratory: <http://www.cevl.msu.edu>

[2] <http://envirosonic.cevl.msu.edu>

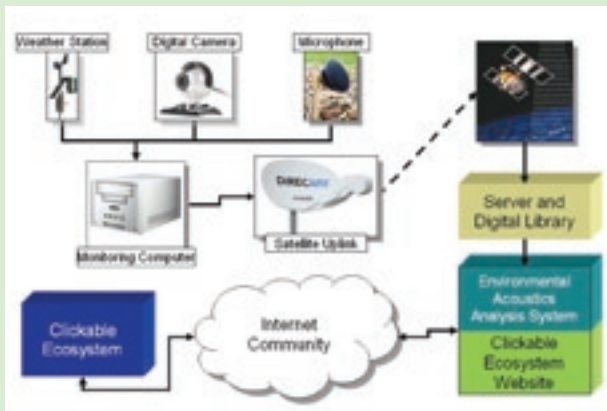


Figure 1. Environmental acoustic monitoring system



Figure 3. A soundscape



Figure 2. Field monitoring equipment

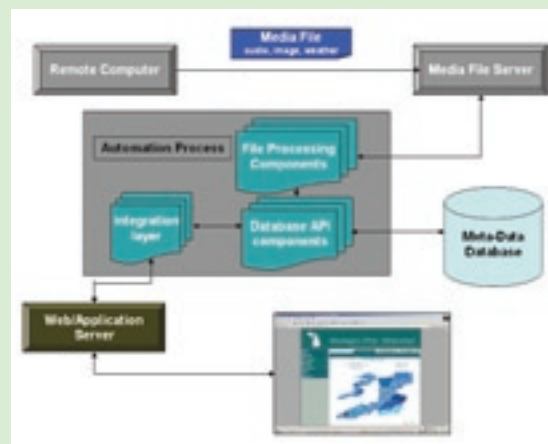


Figure 4. Near real-time acoustic analysis system

8.3.4 Education, Outreach, and Training Recommendations

- **Develop interdisciplinary sensor network curricula, field training, and research programs** to train the next generation of scientists in these emerging tools.
- **Provide outlets and venues for technology exchange** across scientific disciplines.
- **Introduce the utility and application of sensor networks into elementary and secondary education.**
- **Motivate the importance for developing methods of distilling the information collected by sensor networks into forms that are compelling and informative for policy makers and the general public.**
- **Develop standard methods for public reporting of sensor data.**
- **Balance scientific interests with those of public safety agencies** in developing both the sensor network infrastructure as well as enabling technologies.
- **Develop methods to increase awareness of and provide educational resources for A&V tools.**

Appendix A: Workshop Participants

Workshop Organizers

Deborah Estrin
Director, Center for Embedded Networked Sensing (CENS)
Professor, Computer Science Department, UCLA
3531H Boelter Hall
Los Angeles, CA 90095-1596
Ph (310) 206-3923
destrin@CS.UCLA.EDU
<http://cens.ucla.edu/Estrin>

William K. Michener
LTER Network Office
Department of Biology
MSC03 2020
University of New Mexico
Albuquerque, NM 87131-0001
Ph (505) 272-7831
Fax (505) 272-7080
wmichene@lternet.edu

Workshop Steering Committee

David Brady
Electrical and Computer Engineering Department
442 Dana Research Center
Northeastern University
Boston MA 02115
Ph (617) 373-5400
Fax (617) 373-8970
brady@ece.neu.edu
<http://www.cdsp.neu.edu/info/researchgrps/wireless/>

Paul G. Flikkema
College of Engineering and Technology
Northern Arizona University
PO Box 15600
Flagstaff AZ 86011-5600
Ph (928) 523-6114
Fax (928) 523-2300
Paul.Flikkema@nau.edu
<http://www.cet.nau.edu/~pgf/>

Tony R. Fountain
LTER/SDSC Liaison
San Diego Supercomputer Center UCSD
9500 Gilman Drive
La Jolla, CA 92093-0505 USA
Ph (619) 534-8374
Fax (619) 534-5113
fountain@SDSC.EDU
<http://lternet.edu/directory/view.pl?id=tfountain>

Stuart Gage
Michigan State University
Department of Entomology
East Lansing, MI 48824
Ph (517) 355-2135
Fax (517) 353-4354
gages@msu.edu
<http://www.ent.msu.edu/faculty/Gage/>

Scott Matthews
Assistant Professor
Civil and Environmental Engineering and
Engineering and Public Policy
Porter Hall 119
Carnegie Mellon University Pittsburgh, PA 15213-3890
Ph (412) 268-6218
hsm@cmu.edu
<http://www.ce.cmu.edu/~hsm/index.html>

Peter Mikhalevsky
Ocean Sciences Division
Science Applications International Corporation (SAIC)
1710 SAIC Drive (MS T1-3-5)
McLean, VA 22102
Ph (703) 676-4784
Fax (703) 893-8753
peter@osg.saic.com

John Orcutt
Scripps Institution of Oceanography UCSD
8602 La Jolla Shores Drive
La Jolla, CA. 92037
Ph (619) 534-2887
jorcutt@ucsd.edu
<http://roadnet.ucsd.edu/people.html>

Workshop Participants

Payman Arabshahi
California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
MS 238-343
Pasadena, CA 91109 USA
Ph (818) 393-6054
Fax (818) 393-1717
payman@jpl.nasa.gov
<http://dsp.jpl.nasa.gov/members/payman/>

Peter Arzberger
UCSD
9500 Gilman Drive
La Jolla, CA 92093-0043
Ph (858) 822-1079
Fax: (858) 822-4767
parzberg@ucsd.edu
<http://nbcr.ucsd.edu/%7EARzberger.html>

Art Ayres
MariPro, Inc.
1522 Cook Place
Goleta, CA 93117 USA
Ph (805) 879 0109
ayresa@maripro.com

Jon Berger
Institute of Geophysics and Planetary Physics,
Scripps Institution of Oceanography,
University of California San Diego,
La Jolla, CA 92093-0225, USA
Ph (858) 5342889
jberger@ucsd.edu

Gregory Bonito
Graduate Student
Duke University
Department of Biology Box 90338
Durham, NC 27708-0338
Ph (919) 660-7372
gmb2@duke.edu

Mike Botts
ESSC/NSSTC
University of Alabama, Huntsville
Huntsville, AL 35899
Ph (256) 961-7760
Mike.botts@nsstc.uah.edu

Hans-Werner Braun
National Laboratory for Applied Network Research & HPWREN
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
hwb@nlanr.net
hwb@sdsu.edu

Dave Carlson
NCAR/Atmospheric Technology Division
P.O. Box 3000; 1850 Table Mesa Drive
Boulder, CO 80307; USA
Ph (303) 497-8833
Fax (303) 497-8770
dcarlson@atd.ucar.edu
<http://www.atd.ucar.edu/>

Neil Cobb
Northern Arizona University
Associate Director Merriam-Powell Center for Environmental
Research
Hanley Hall
Flagstaff, AZ 86011
Ph (928) 523-5528
Fax (928) 523-7500
Neil.Cobb@nau.edu

Dennis Conlon
Office of Naval Research
800 N. Guinex St
Arlington, VA 27217
Ph (263) 696-4720
Fax (263) 696-2710
conlond@onr.navy.mil

David Culler
Computer Science Division #1776
627 Soda Hall
University of California, Berkeley
Berkeley, CA 94720-1776
Ph (510) 643-7572
Fax (510) 643-7352
culler@cs.berkeley.edu
<http://www.cs.berkeley.edu/~culler/>

Sanjoy Dasgupta
Assistant Professor
Department of Computer Science and Engineering, UCSD
University of California, San Diego
9500 Gilman Drive, Dept. 0114
La Jolla, CA 92093-0114
Ph (858) 822-5270
dasgupta@cs.ucsd.edu

Jennifer Doherty
Graduate Student
Department of Organismic Ecology and Evolution
University of California, Los Angeles
Box 951606 LA, CA 9005-1606
dohertyj@ucla.edu

Wynn Eberhard
NOAA Environmental Technology Laboratory
325 Broadway
Boulder, CO 80305
Ph (303) 497-6560
Fax (303) 497-5318
wynn.eberhard@noaa.gov

Jeremy Elson
Research Staff
Center for Embedded Networked Sensing (CENS)
University of California Los Angeles
Department of Computer Science
3440 Boelter Hall
Los Angeles CA 90095
Ph (310) 206-3925
jelson@cs.ucla.edu

John Fisher
MIT CSAIL
200 Technology Square
NE43-V 626
Cambridge, MA 02139
Ph (617) 253-0788
Fax (617) 258-6287
fisher@csail.mit.edu
<http://www.ai.mit.edu/people/fisher/>

Ed Frieman
Scripps Institute of Oceanography, UCSD
8602 La Jolla Shores Dr
La Jolla, CA 92037
(858) 826-6756
efrieman@ucsd.edu

David Fries
Center for Ocean Technology
140 7th Avenue South
St. Petersburg, FL 33701-5016
University of South Florida
Ph (727) 553-3961
Fax (727) 553-3967
dfries@marine.usf.edu

Eric Frost
Immersive Visualization Center
San Diego State University
CAL-(IT)2
San Diego, CA 92182
Ph (619) 594-5003
Fax (619) 594-4272
Eric.frost@sdsu.edu
<http://map.sdsu.edu/visual>

John Gamon
Department of Biology & Microbiology
Biological Sciences
5151 State University Drive
California State University, LA
Los Angeles, California 90032-8201
Ph (323) 343-2066
Fax (323) 343-6451
jgamon@calstatela.edu
<http://web.calstatela.edu/faculty/jgamon/jgamon.htm>

Lewis Girod
UCLA/LECS Laboratory
420 Westwood Plaza
3731 Boelter Hall
LA CA 90095 USA
Ph (310) 206-3925
Fax (501) 421-4987
girod@cs.ucla.edu
<http://lecs.cs.ucla.edu/~girod/official/>

Jeffrey Goldman
Infrastructure for Biology at Regional to Continental Scales
American Institute of Biological Sciences
1444 Eye Street, NW, Suite 200, Washington, DC 20005
Ph (202) 628-1500 x225
Fax (202) 628-1509
jgoldman@aibs.org

Michael Hamilton
University of California
James Reserve Director
PO Box 1775
Idyllwild, CA 92549
Ph (909) 659-3811
Fax (909) 659-0553
director@jamesreserve.edu
<http://www.jamesreserve.edu>

Mark Hansen
Associate Professor
Department of Statistics
University of California, Los Angeles
6119 Mathematical Sciences Building
Los Angeles, CA 90095-1554
Ph (310) 206-8375
Fax (310) 206-5658
cocteau@stat.ucla.edu
<http://www.stat.ucla.edu/~cocteau/>

Paul C. Hanson
UW-Madison
Center for Limnology
680 North Park Street
Madison WI 53706-1492
Ph (608) 262-5953
pchanson@wisc.edu
<http://limnology.wisc.edu/personnel/hanson/hanson.html>

Tom Harmon
School of Engineering
UC Merced
PO Box 2039
Merced, CA 95344
Ph (209) 734-4337
tharmon@ucmerced.edu
<http://www.cee.ucla.edu/faculty/harmon.htm>

Paul Havinga
University of Twente
Department of Computer Science
P.O. Box 217
7500 AE Enschede
the Netherlands
Ph: +31 53 4894619
Fax: +31 53 4894590
havinga@cs.utwente.nl
<http://www.cs.utwente.nl/~havinga>

John Heidemann
USC/Information Sciences Institute
Suite 1001
4676 Admiralty Way
Marina Del Rey, CA 90292-6695
Ph (310) 448-8708
johnh@isi.edu
<http://www.isi.edu/~johnh/index.html>

John Helly
UCSD/SDSC
9500 Gilman Drive
La Jolla, CA 92093-0505
Ph 858 534 5060
hellyj@ucsd.edu

Masayuki Hirafuji
Computational Modeling Lab.
NARC Tsukuba 305-8666 Japan
Ph: +81-298-38-7177
Fax: +81-298-38-8551
hirafuji@affrc.go.jp
<http://model.job.affrc.go.jp>

Mike Horton
Crossbow Technology, Inc.
41 E. Daggett Dr.
San Jose, CA 95134
Ph (408) 965-3300
Fax (408) 324-4840
mhorton@xbow.com

Bill Kaiser
Electrical Engineering Department
56-125B Engineering IV Building
Box 951594
University of California, Los Angeles
Los Angeles, CA 90095-1594
Ph (310) 825-2647
Fax (310) 206-4833
kaiser@ee.ucla.edu
<http://www.ee.ucla.edu/faculty/bios/kaiser.htm>

Josh Karlin
Graduate Student
Computer Science Department
MSC01 1130
1 University of New Mexico
Albuquerque, NM 87131-0001
karlinjf@cs.unm.edu

Brad Karp
Staff Researcher / Intel Research
Computer Science Department
Carnegie Mellon University
417 South Craig St.
Suite 300
Pittsburgh, PA 15213
Ph (412) 605-1209
Fax (412) 605-1306
bkarp@cs.cmu.edu or bkarp@intel-research.net
<http://www-2.cs.cmu.edu/~bkarp/>

John Kim
Field Station Programs
College of Sciences
San Diego State University
5500 Campanile Dr
San Diego, CA 92182-4614
Ph (619) 594-5386
jkim@sciences.sdsu.edu

Barbara Kimbell
Special Assistant to the Vice Provost for Research
University of New Mexico
801 University Blvd SE Suite 301
Albuquerque, NM 87107
Ph (505) 272-7099
bkimbell@unm.edu

George Koch
Associate Professor
Department of Biological Sciences
Northern Arizona University
NAU Box 5640
Flagstaff, AZ 86011-5640
Ph (928) 523-7150
George.Koch@nau.edu

Tim Kratz
Trout Lake Station
University of Wisconsin-Madison
10810 Cty Hwy N
Boulder Junction, WI USA 54512-9733
Ph (715) 356-9494
Fax (715) 356-6866
tkkratz@wisc.edu

Ron Lake
Galdos Systems, Inc.
Suite 200
1155 West Pender Street
Vancouver, B.C. V6E 2P4
Canada
Ph (604) 484-2750
rlake@galdosinc.com

Phil Levis
Graduate Student
Computer Science Division #1776
467 Soda Hall
University of California, Berkeley
Berkeley, CA 94720-1776
Ph (510) 290-5283
Fax (510) 643-7352
pal@eecs.berkeley.edu
<http://www.eecs.berkeley.edu/~pal/>

Alex Lightman
Visiting Scholar
Cal-(IT)2, SDSU
Cal-(IT)2 Director's Office
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093
Ph (310) 717-7745
Lxlightman@aol.com

Fang Pang Lin
Research Scientist
National Center for High-Performance Computing, Taiwan
7, R&D Rd. VI
Science-based Industrial Park
Hsinchu, Taiwan, R.O.C.
Ph 886-3-5776085
Fax 886-3-5776082
fplin@nchc.gov.tw
www.nchc.gov.tw

Jessica Lundquist
Graduate Student
Scripps Institution of Oceanography UCSD
8602 La Jolla Shores Drive
La Jolla, CA. 92037
Ph: (858) 534-1504
jessica@coast.ucsd.edu
http://meteora.ucsd.edu/cap/snow_monitor.html

Tim Lyons
Univ. of Missouri
101 Geological Sciences Building
Columbia, MO 65211, USA
Fax: (573) 882 5458
Tel: (573) 882-6328
LyonsT@missouri.edu
<http://web.missouri.edu/~geolwww/faculty/lyons.html>

Arthur Maccabe
Department of Computer Science
MSC01 1130
University of New Mexico
Albuquerque, NM 87131-0001
Ph (505) 277-6504
Fax (505) 277-6927
maccabe@cs.unm.edu
<http://www.cs.unm.edu/~maccabe/>

James Moore
UCAR, Joint Office for Science Support
P.O. Box 3000
Boulder, CO 80307-3000
Ph 303-497-8635
Fax 303-497-8158
jmoore@ucar.edu
<http://www.joss.ucar.edu>

Rob Nowak
ECE Department
1415 Engineering Drive
Madison, WI 53706
nowak@engr.wisc.edu
www.ece.wisc.edu/~nowak

Walter C. Oechel
Professor of Biology and Director
Global Change Research Group
San Diego State University
San Diego, CA 92182
(619) 594-6613
oechel@sunstroke.sdsu.edu
<http://www.sci.sdsu.edu/GCRG>

Clayton Okino
Jet Propulsion Laboratory
4800 Oak Grove Drive
MS 238-343
Pasadena, CA 91109 USA
Ph (818) 393-6668
Fax (818) 393-1717
cokino@jpl.nasa.gov
<http://dsp.jpl.nasa.gov/members/clay/>

Eric Osterweil
Center for Embedded Networked Sensing (CENS)
University of California Los Angeles
Department of Computer Science
3440 Boelter Hall
Los Angeles CA 90095
eoster@CENS.UCLA.EDU

Raju Pandey
Department of Computer Science
3041 Engineering Unit II
One Shields Avenue
University of California, Davis, CA 95616
Ph (530) 752-3584
Fax (530) 752-4767
pandey@cs.ucdavis.edu
<http://pdclab.cs.ucdavis.edu/~pandey/>

Adrian Perrig
ECE - CMU
Hamerschlag Hall
5000 Forbes Avenue
Pittsburgh PA 15213
Ph (412) 268 2242
perrig@cmu.edu
<http://www.ece.cmu.edu/~adrian/home.html>

Philip Papadopoulos
UC San Diego
San Diego Supercomputer Center, MC 0505
9500 Gilman Drive
La Jolla CA 92093-0505
Ph (858) 822-3628
Fax (858) 534-5152
phil@sdsc.edu
<http://www.sdsc.edu/Visitors/contact.html>

Philip Rundel
Department of Organismic Biology, Ecology, and Evolution
University of California, Los Angeles
Los Angeles, CA 90095-1606
Ph (310) 825-8777
Fax (310) 825-9433
rundel@biology.ucla.edu
<http://research.mednet.ucla.edu/cfm/lifesci/OBEEfacultyindiv.cfm?FacultyKey=1131>

William H. Sanders
University of Illinois
212 Coordinated Science Laboratory, MC-228
1308 West Main Street
Urbana, IL 61801-2307
Ph (217) 333-0345
Fax: (217) 244-3359
whs@crhc.uiuc.edu
<http://www.crhc.uiuc.edu/Faculty/whs.html>

Art Sanderson
Department of Electrical
7015 Low Center for Industrial Inn
Rensselaer Polytechnic Institute
110 8th St.
Troy, NY 12180-3590
Ph (518) 276-4873
sandea@rpi.edu

Dogan Seber
Geoinformatics Lead
San Diego Supercomputer Center
9500 Gilman Drive
MC 0505
La Jolla, CA 92093
(858) 822-5409
seber@sdsc.edu
<http://atlas.geo.cornell.edu/people/seber.html>

Frieder Seible
Dean of Jacobs School Of Engineering
UCSD
Mail Code 0085/Bldg 409
La Jolla, CA 92093
Ph (858) 534-4640
Fax (858) 534-6373
fseible@ucsd.edu
<http://www.structures.ucsd.edu/Faculty/Seible.shtml>

Srini Seshan
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213-3891
Ph (412) 268-8734
Fax (412) 268-5576
srini@cmu.edu
<http://www-2.cs.cmu.edu/~srini/>

Sedra Shapiro
Field Station Programs
College of Sciences
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-4614
Ph (619) 594-5386
sshapiro@sciences.sdsu.edu
<http://www.sci.sdsu.edu/BFS/>

Roy Shea
UCLA
Department of Computer Science
3440 Boelter Hall
Los Angeles, CA 90095
Ph (310) 824-1405
roy@CS.UCLA.EDU

Shinji Shimojo
Cybermedia Center
Osaka University
5-1 Mihogaoka, IBARAKI
Osaka 567-0047 JAPAN
Ph +81-6-6879-8790
Fax +81-6-6879-8794
shimojo@cmc.osaka-u.ac.jp
<http://www.ais.cmc.osaka-u.ac.jp/~shimojo/mainE.html>

David Skole
Department of Geography
Michigan State University
East Lansing, MI 48824
Ph (517) 432-7774
Fax (517) 353-2932
skole@msu.edu or skole@pilot.msu.edu

Larry Smarr
Cal-(IT)² Director's Office
University of California,
San Diego
9500 Gilman Drive
La Jolla, CA 92093-0405
Ph (858) 822-1189
Fax (858) 822-3912
lsmarr@ucsd.edu
<http://www.jacobsschool.ucsd.edu/~lsmarr/>

John Stankovic
Department of Computer Science
School of Engineering and Applied Science
University of Virginia
151 Engineer's Way, P.O. Box 400740
Charlottesville, VA 22904-4740
Ph (434) 982-2275
Fax (434) 982-2214
stankovic@cs.virginia.edu
<http://www.cs.virginia.edu/brochure/profs/stankovic.html>

Robert Stevenson
Associate Professor
Department of Biology
University of Massachusetts, Boston
Boston, Massachusetts 02125
Ph (617) 282-6572
Fax (617) 287-6650
robert.Stevenson@umb.edu
http://www.bio.umb.edu/WhosWho/Faculty/swifty_cv.html

Robert Szewczyk
Graduate Student
University of California at Berkeley
Computer Science Division
467 Soda Hall
Berkeley, CA 94720
Ph (510) 387-0532
szewczyk@cs.berkeley.edu

Mohan Trivedi
Professor of Electrical and Computer Engineering, UCSD
UC San Diego, CVRR
9500 Gilman Drive 0434
La Jolla, CA 92093-0434
Phone: 858-822-0075
Fax 858-534-1004
trivedi@ece.ucsd.edu

Ahmad Varoqua
Field Station Programs
College of Sciences
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-4614
Ph (619) 594-8026
avaroqua@sciences.sdsu.edu

Frank Vernon
Associate Research Geophysicist
Scripps Institution of Oceanography UCSD
8602 La Jolla Shores Drive
La Jolla, CA. 92037
flvernon@ucsd.edu

Hanbiao Wang
Graduate Student
UCLA Computer Science Department
3440 Boelter Hall
Los Angeles, CA 90095-1596
hbwang@lecs.cs.ucla.edu

Michael Wimbrow
UC Riverside – James Reserve
19412 Dorado Drive
Trabuco Canyon, CA 92679
Ph (909) 659-0024
mike.wimbrow@jamesreserve.edu

Fan Ye
Graduate Student
UCLA Computer Science Department
4805 Boelter Hall
Los Angeles, CA 90095-1596
Ph (310) 825 4838
Fax: 310 825 2273
yefan@cs.ucla.edu
<http://www.cs.ucla.edu/~yefan/>

NSF REPRESENTATIVES

Rachael Craig
Division of Earth Sciences
National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230
Ph (703) 292-8233
Fax (703) 292-9025
rcraig@nsf.gov

Dylan George
National Science Foundation
Division of Environmental Biology - NEON
4201 Wilson Boulevard
Arlington, Virginia 22230
Ph (703) 292-8480
Fax (703) 292-9064
dgeorge@nsf.gov

Alexandra Isern
Division of Ocean Sciences
Program Director Ocean Technology
National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230
Ph (703) 292-8583
aisern@nsf.gov

Stephen Meacham
Division of Atmospheric Sciences- ITR Program Director
National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230
Ph (703) 292-8527
Fax (703) 292-9022
smeacham@nsf.gov

Priscilla Nelson
Directorate for Engineering - Senior Advisor
National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230
Ph (703) 292-7018
Fax (703) 292-9013
pnelson@nsf.gov
www.eng.nsf.gov

Louie Tupas
Office of Polar Programs
National Science Foundation
4201 Wilson Boulevard
Arlington Virginia 22230
Ph (703) 292-8092
Fax (703) 292-9082
ltupas@nsf.gov

www.LTERNET.edu/sensor_report

Sponsored by the National Science Foundation through an award to the University of New Mexico's Long Term Ecological Research Network Office. Additional support provided by the Center for Embedded Networked Sensing.



The University of New Mexico

