Cyberinfrastructure for Environmental Observations, Analysis, and Forecasting: A Cyberinformatics Forum

Workshop Report

May 5–7, 2008

National Center for Atmospheric Research

Boulder, Colorado

Report prepared by

Peter Backlund, National Center for Atmospheric Research

Peter Fox, Rensselaer Polytechnic Institute

Katherine Lawrence, University of Michigan

Greg Guibert, National Center for Atmospheric Research

Introduction

In May 2008, the National Center for Atmospheric Research (NCAR) hosted a workshop on the creation and use of new cyberinfrastructure (CI) capabilities to support environmental research and education. The goal of the "Cyberinfrastructure for Environmental Observations, Analysis, and Forecasting: A Cyberinformatics Forum" was to assist the National Science Foundation (NSF) in considering how it might most effectively craft CI programs to support a broad range of environmental science and education goals over the next decade. The workshop was intended as an important continuing step in identifying the key issues and challenges in the development and sustainability of cyberinfrastructure for environmental observatories as well as exploring the means of addressing them. The objective was not to try and build a consensus for a fully defined proposal or plan, but rather to identify opportunities through stimulating a constructive dialogue among scientists and educators from different environmental communities, and between environmental scientists and information technologists.

This three-day workshop was one of a series that have been held as part of an NSF-supported project on coordination of cyberinfrastructure for environmental observation. The project has organized meetings dedicated to cyberinfrastructure for Arctic and Ocean observations, as well as several devoted to issues that cut across focused observational activities and programs. This workshop also followed on to and benefitted from an NSF-hosted meeting in early 2008 that examined the concept of a Federation of Environmental Observation Networks (FEON). Approximately 70 researchers, educators, government agency representatives, and technologists from across the United States participated (see Appendix A for full participant list). This report is the product of the workshop. The intent was to examine opportunities for applying cyberinfrastructure in environmental research and education, identify significant issues and challenges, and provide a roadmap for addressing key questions. This report will be submitted to NSF, published on the web, and presented at future scientific and technical conferences.

Methodology

Participants

Prior to arrival, workshop participants submitted a position statement addressing three questions:

1. What are the exciting opportunities in cyberinfrastructure for environmental observatories?
2. What recent technical innovation do you see as most important?
3. Please describe the non-technical factors that you see as most important for successful cyberinfrastructure implementation.

See Appendix B for a sample of these comments. Responses to the position statements served to create broad framework for the workshop.

Seventy people from across the United States participated in the workshop, bringing together academics, practitioners, technologies, and educators. Forty-three were from environmentally oriented academic departments, research centers, or facilities, including 12 from NCAR/UCAR, representing a wide range of scientific disciplines such as:

- Arctic science
- Atmospheric science
- Biomedical informatics
- Civil and environmental engineering, hydrology
- Climate change, both regional and global
- Environmental science
- Geological science , geography, cartography
- Geospatial Science, GIS
- Ocean/Marine Sciences
- Paleoclimatology and Paleoceanography
- Phenology Networks
- Plant ecology
- Polar research
- Solar science

Sixteen were from computer science–oriented academic departments or research centers, many with specific interests in cyberinfrastructure and high-performance computing, information science, informatics, and information management. There were five representatives from government agencies and four participants focused on education and outreach, representing museums, libraries, and public schools. Two participants were social scientists.

In addition to these disciplinary interests, participants described a range of other topical interests, including:

- Collaboration
- Knowledge representation and reasoning
- Cyberinfrastructure: archaeological data, ocean observatories, ecological observatories, earth-observing platforms, sensors, real-time data systems, analysis and display, scientific workflows
- Data: management, infrastructure, design, development, evaluation, use, integration, coordination, archiving, access, visualization, standards, quality, metadata, civilian science, web services
- Real-time data: instrumentation, field observations, distribution, data services, numerical weather prediction, datasets, 4-D assessment
- Modeling: climate, agriculture affected by weather, surface dynamics
- Interoperability standards: marine science, semantic, spatial, geospatial data
- Observatory/monitoring systems: ocean, coastal, environment, biodiversity
- Applications and extensions of scientific research: science curriculum, museums, immersive theaters, societal benefit

Workshop Structure

The workshop featured a mix of plenary presentations and breakout discussion groups (see Appendix C for a complete agenda). The opening plenary presentation detailed the key issues that participants had previously identified in their submitted position statements. Other presentations followed and included Barbara Minsker discussing the initial activities of FEON (Federation of Environmental Observatory Networks); Deborah McGuinness on semantic e-Science; NEON (National Ecological Observatory Network) CEO Dave Schimel; and, lessons learned from NASA's earth science data systems, presented by Hampapuram Ramapriyan. A forward-looking wrap-up by Sara Graves of the University of Alabama in Huntsville concluded the workshop.

The breakout groups were held on the first and second day of the workshop and were intended to crosscut the issues of cyberinfrastructure for environmental observations, analysis, and forecasting in two ways. The first set of breakout groups focused on three overarching challenges that had been identified as central issues during the FEON cyberinfrastructure coordination meeting:

- Organization
- Standards
- Technology

The groups were seeded with a selection of questions and issues to stimulate discussion (See Appendix D for a sample of topics).

The second set of breakout groups discussed three complementary issues that cut across the breakout topics of the first day:

- Instruments and measurements
- Data processing, assimilation, and archiving
- Information distribution and accessibility

During these discussions, the groups were encouraged to identify gaps, determine readiness and capability, and consider near-term versus long-term needs, particularly those could be enhanced by increased coordination. The groups were also prompted by a smaller set of framing questions (also listed in Appendix D).

Each breakout group had a facilitator and a scribe. Written breakout group notes were reviewed and coded to identify key themes and are summarized in this report.

Breakout Groups and Discussions: Key Themes

The breakout groups discussed three overarching challenges to environmental observatories (EOs) in general and a federated network of EOs more specifically. The first three groups addressed organization, standards, and technology. The second three breakout groups were topics chosen as cross-cutting issues: instruments and measurements; data processing, assimilation, and archiving; and, information distribution and accessibility. The group discussions for each of these six topics are summarized below.

Organization

The organizational issues associated with building and maintaining cyberobservatories are complex and there are a number of pre-existing challenges that require meaningful incentives and pioneering solutions in order to overcome them.

Underexposed tensions hindering participation

Current resource and recognition tensions among potential collaborating EOs hinder current cyberinfrastructure efforts and will likely restrict future growth unless explicitly addressed. The availability of personnel and surplus IT resources for collaborative projects are central issues. Questions and concerns about continuity and sustainability of funding for developed systems as a tradeoff for new systems also add additional uncertainty and apprehension for collaborators. Institutional and disciplinary reward and recognition structures can adversely affect individual motivation and prioritization of collaborative endeavors. Finally, there was concern that data standardization could reduce individual creativity in the scientific process.

Impact of academic infrastructure on CI efforts

Academic infrastructure presents specific challenges for attracting collaborators to work on CI. As was widely noted, the implementation of solutions for real-world problems is not well rewarded during the tenure review process. As a consequence, faculty have greater incentive to produce research-quality systems rather than sustain CI ventures to maturity. Within individual disciplines, scientists may also be competitive across institutions and departments and worry about being “scooped” by colleagues. Because few universities have established programs to train students or new scientists in the multi-disciplinary knowledge that bridges CI and science, insufficient personnel are knowledgeable about the complexities of supporting data management for large research endeavors.

What are NSF’s funding goals?

Clearer signals about CI funding priorities and goals from NSF would encourage investment and commitment from collaborators. The incentivizing and structuring of CI efforts are dependent on overarching organizational objectives from funders. Without articulated mandates supported by funding, uncertainty will continue to restrict growth potential.

Incentives for collaboration

Despite uncertainties and hurdles, there exist a number of incentives for participation in cyberobservatories. For example, collaborators may discover that funding can be used more effectively, and in the process learn best practices, when pursuing common objectives with likeminded colleagues. The ease of data sharing among centers and observatories may also become a powerful attraction. Participants suggested that customized sections targeted toward meeting the requirements of specific groups would likely promote early adoption. Grant guidelines could also explicitly increase participation by specifically requiring data sharing or outreach to potential users.

FEON as a possible solution

Participants discussed how a united observatory network such as FEON (Federated Environmental Observatory Network) could be most successful. FEON would need clear structures and lines of authority, supplemented by an advisory board to represent stakeholders’ interests. To function as source of centralized standards, it would also require connectivity with other agencies, data centers, and research institutions. Full interoperability among observatories, however, may require a broader public policy commitment. Partnerships with industry and leveraging private sector resources would be an overall benefit largely unrealized within individual EOs.

A good starting place for achieving successful development and outreach would be use cases, and as the system grows, it might support a "virtual organization in a box" to help its constituent members avoid solving the same issues repeatedly.

Bridging across groups & disciplines

EOs will need to bridge multiple scientific and technological communities for both development and robust ongoing participation. All interested parties will need to communicate common objectives and values while recognizing that different systems and endeavors may be at varying levels of development. For example, the creation of interoperability through standardized metadata and services alone requires the blending of disciplines and specialties—domain sciences, computer science and information technology, semantics, ontologies, cognition, and high performance computing. Moreover, there is hope that EOs will promote better access through outreach, particularly beyond the "science only" use of CI to make the resources available to those in educational, social science, and policy/legislative areas.

Designing effective systems for all participants

CI presents some acute usability challenges. By identifying multiple user audiences, exploring their usage needs, and developing use cases, EOs will improve accessibility to new constituents and research communities. Some of the features necessary for wider adoption include integrated data repositories, interoperability to make data sharing easier, stable technology, and sustained human technical support.

Standards

The development, adoption, and interoperability of standards represent significant ongoing challenges for EO networking and collaboration. There are, however, important lessons that have been learned from prior experiences as well as a keen awareness of what remains to be better understood.

Development of standards

EOs may eventually require dozens of standards governing areas of CI including data exchange, data encodings, ontologies, provenance and other metadata, and representations of space and time, among many others. Standards are more effective when they are purpose driven, focusing on high-profile scenarios and the use cases associated with them (e.g., disaster response, climate change). Likewise, agency or industry requirements may prove useful for motivating the establishment of standards. However, prior experience suggests that standards are more effective when evolved from a community of practice (vs. top-down imposition), as illustrated by NASA's Earth

Observing System (EOS) program[1]. One way of drawing on the experiences and best practices of developers and users is the creation of a repository of shared experiences (such as the Open Ontology Repository[2]), which would include use cases.

Ontologies

Ontologies—the formal representation of concepts in a domain and the relationship between those concepts—were a particular topic of interest within the standards breakout group. They suggested annotating ontologies to make them usable across disciplines while avoiding numerous terms for the same concept. Also they discussed the importance of maintaining compatibility across systems, without "showing the guts of the system".

Adoption of standards

Just as important as the effective development of standards is the willingness of the communities' developers and members to accept and support those standards. An entity like FEON could serve a coordinating role, particularly to review standards, convey the scope of the standards, suggest appropriate usages, and demonstrate how others are using them. NASA's Standards Process Group[3] offers a successful example of facilitating the identification and distribution of standards. Likewise, the Open Geospacial Consortium (OGC) Interoperability Program[4] has an approach for developing community data models that might provide useful examples.

Interoperability

The topic of interoperability is closely related to standards. However, before more work can be done on interoperability, EOs need to build awareness of and incentives to use existing standards. Meanwhile, an organization such as FEON could help coordinate focused, technical discussions among EOs (nationally and internationally). There is some concern that many funded projects are not scoped to consider interoperability and thus may not be adequately connected; funded positions within FEON could be alert to opportunities for interoperability with other projects and communities.

Barriers

Despite awareness of best practices in the development of standards, there remains a lack of understanding how communities react to new standards. Unfortunately, standards development is not considered "science" by funding agencies so it is often neglected by academics. It is clear, however, that many people are reluctant to adopt new technologies, often due to problems of usability and accessibility. Furthermore, developers of standards need to think beyond the most obvious audiences (e.g., research, policy) and consider how to make those standards accessible to K–12 users.

---

[1] NASA Earth Observing System: http://eospso.gsfc.nasa.gov/

[2] Open Ontology Repository: http://oor-01.cim3.net/ For more information see: http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008_Communique#nid1GUJ

[3] NASA Standards Process Group: http://www.esdswg.org/spg/

[4] OGC Interoperability Program: http://www.opengeospatial.org/ogc/programs/ip

Technology

The technology breakout group focused primarily on data issues, even when discussions incorporated related topics such as models, access, security, and architecture.

Data

Data reception was the technology topic that dominated the discussion due, in large part, to the breadth of issues associated with scalability, transport, security, metadata, quality control, provenance, semantic and contextual information, and adaptability or flexibility. For example, as one participant pointed out, metadata is often inadequately documented, so it may diverge over time. Participants noted that bringing new data sources into a system can be time consuming.

Data storage was another key issue. There is little consensus on storage best practice, which generated significant debate about whether data should be stored in a few large locations or many small ones. Likewise, questions about how data will be captured and maintained represent important challenges to be addressed. Storage is not a "one-way" process because data from sensors may require post-storage calibration. In such cases, experts do not yet have enough information to evaluate the tradeoffs of keeping primary data versus only recomputed data versus discarding some data. Another question is whether data could be offered as a service, rather than as files. For example, URLs would provide access to the latest, best version of the data.

Models

Participants identified some key opportunities for using models with EOs. For example, they saw an opportunity to use real-time data in models to help guide sensors, as well as the utility of viewing or analyzing a wider range of data through the EO infrastructure. However, a difficulty is that models are not always apparent to those who are less familiar with the representation or algorithms—even informed scientists may struggle with them—so the original algorithms or code has to be retained along with data.

Access

Access to data, models, and output is a central issue. One aspiration is to make archived data and model output available more broadly, but often non-experts find the access interfaces too difficult to navigate. Google Earth is a tool that offers greater public reach but has some technical and accuracy limitations that would need to be addressed through the provision of accurate, validated datasets. Also, given the anticipated volumes of data produced by EOs, how long data needs to be kept "active" or available "on demand" given the resources necessary to maintain it will be an ongoing concern. For some environmental topics, such as climate research, scientists need 25 years or more of historic data. To have control over these types of decisions, some may decide that environmental science should have a dedicated system for environmental prediction and analysis.

Security

As EOs broaden access, the security of the system will depend on authorizations, which in turn will require coordination across institutions, and possibly with other countries. Some aspects of data security are of less of a concern than others, but prominent worries include methods of controlling hacking and the bogus submission of data. Metrics, audit trails, and backups are a few ways of addressing these issues.

Architecture/CI

The particular challenge of architecture and CI design for EOs is balancing the competing forces that are converging technologies for interoperability, while allowing for individual evolution in response to changing needs, instruments, and formats without compromising performance. Experts still need to determine how to harness grids and virtualization technologies to provide resources on demand. Layered on top of these technical issues is the question of how to structure the virtual organizations that manage these systems. For example, grid computing may be too heterogeneous for variable demand processes such as forecasting, yet relying on outsourced cloud computing (e.g., virtualization on Amazon or Google) may be insufficiently safe and effective. More detailed exploration of these topics was deemed beyond the scope of the workshop.

Instruments and Measurements

This group focused on data and sensors, raising the costs and benefits associated with using "commodity," off-the-shelf technologies rather than custom equipment and the participation of citizen scientist that such technologies would allow. Cross-cutting themes included: moving from research to operations, incentives and disincentives to use particular technologies, engineering systems-level solutions that incorporate adequate standards, and quality.

Data: Capturing data in new ways

To date, several competing approaches exist for capturing and providing data, each with its own challenges. One is to offer quick delivery in near real-time. Another is to process the data into "information" for later use. Yet another is to merge individual instruments' data in a catalog system, and a fourth is to offer streams of data based on specific topics. A popular "Web 2.0" way of working with non-scientific data is mashups—pulling data from multiple sources into a single, integrated tool—but they are not yet common in science applications, possibly because the repurposing of unintended data sources requires strong metadata.

Sensors

To build new sensors, scientists have several, non-exclusive options. They can draw on existing, off-the-shelf technologies; they can make do with research prototypes that may be more customized but less proven; and, they can draw on lessons learned from other, more experienced disciplines (e.g., computer science, military applications). To achieve the goals of real-time and dynamic sensing, EOs require autonomous machine-to-machine decision making, good communications, calibration, quality control, error handling, and sufficient manpower to run the systems.

For any EO sensor contributing to a network, researchers will need to incorporate metadata into the sensor design. Such metadata could include changes, calibration, and configuration parameters. Sensor design and configuration for metadata would be predicated on a strong and accepted set of standards. It is difficult to incorporate this level of sensor sophistication into legacy sensors systems.

Commodities: Off-the-shelf technology

"Commodity" technologies may offer attractive cost saving promise for EOs that will have to be weighed against some clear pitfalls. Instruments and sensors, data storage, computational power, and equipment become commodities when there is the demand, infrastructure, and advanced development to make them inexpensive. For example, Amazon and Google have made data storage a commodity. Already sufficient demand exists in some earth sensing systems, such as weather, which has customer base of farmers, sailors, pilots, hobbyists. While growing demand and maturity will reduce costs and increase capacity over time, mass production will also drive a need for automated quality control. Likewise, plug-and-play is not guaranteed to follow accepted scientific standards (such as communication, calibration, self-reporting), and therefore may not ultimately result in cost savings.

Citizen scientists

The benefit of cheap off-the-shelf technologies is the greater participation of citizen scientists. They may join EO efforts through active recruitment, volunteering, receiving pay for "voluntary" contributions, and industry partnerships to encourage citizen or employee participation. With this broader participation come additional challenges and issues, including quality control, privacy concerns, and translation of data from one scale to another. Also, EOs will need to consider what incentives citizen scientists have to participate—feedback, recognition, real-time information—and whether meeting those needs will be consistent with larger mission objectives.

FEON

The group closed with a brief description of the viability of an EO federation such as FEON. Some agreed that FEON, as currently envisioned, would be good as a bottom-up model of collaboration, but that it would have little financing or inherent power to influence and drive progress. Such an organization would need investment, leadership, incentives for contributing, rewards for implementing standards, and a short-term payback. Some of the barriers to buy-in include a need for sensor systems, a perceived loss of custom technology that would be more competitive than commodity technologies, and the effort necessary to make collaboration successful.

Data Processing, Assimilation, and Archiving

This group explored the role that a FEON-like organization could offer as a way of adopting and sharing best practices for the bridging of raw data and its subsequent use, specifically in the areas of data preservation, virtualization, metadata, security policies, and content management. The group sought to identify how they could share best practices while retaining the necessary flexibility to customize for the needs of specific environmental observatory communities. They also briefly discussed the specific topics of virtualization and automation.

Role of an organization like FEON

An EO federation could be a source of best practices by vetting existing methods, processes, and techniques. It could offer guidelines for budgeting, planning, and cost projections while collaborating with agencies to establish standard protocols and metrics. Some of the standards that would need to be coordinated might include data, metadata, query interface, communication protocols, federated identity and security, and level of service (e.g., shelf life of data). A federation could provide a common language for establishing and expressing data policies along with recommendations for metadata that is consistent with a given data policy.

The federation could also support the identification and/or development of effective content management systems, resources, and tools, providing a unified voice for the advocacy of customizations to support EOs. The consolidated efforts of a federation would advance academic systems, too, by advocating for incentives for sharing, offering credit for contributions, and educating students about best practices.

Virtualization

Participants suggested that virtualization may be useful in some areas such as data service, but mentioned it was likely too early for computational models to use it. Virtualization also raises interesting issues around security and migrating complex services to commodity service providers.

Workflows and automation

Workflows offer promising ways of automating many of the major time- or labor-intensive functions of EOs. Beginning with data, workflows could facilitate data submission, ingest, and metadata tagging—though standards would be required to share the data across EOs—and subsequently they could allow simpler data mining across EOs. They would serve as a gateway for identifying model types, applications, and parameters to make it easier to locate them. Workflows offer a paradigm for setting up automated analysis, virtualization, curation, and preservation.

Information Distribution and Accessibility

This group's discussion was structured by four main questions:

- How do we enable interdisciplinary use, especially for other audiences?
- To what extent do we need use cases to enable interdisciplinary use?
- How do we determine the existence and engagement of stakeholders (in a sustained way)?
- How do we understand market forces, incentives and disincentives, buy-in factors, acceptance, trust, priorities, needs, and social and cognitive factors?

Search

Effective information searching remains challenging. Because datasets are often difficult to find and researchers are frequently unaware of interdisciplinary resources such as the NCAR Community Data Portal[5], they often turn first Google or ask a librarian for assistance. Google searches present several problems for accessing relevant data: simple searches do not have sufficient focus, nor do they find registered resources; retrieval rankings are constantly changing and can be advertising driven; and, information recall is poor and may not accurately reflect the user requirements. Participants discussed the possibility of using a tool like FriendFeed[6] or del.icio.us[7] to share with colleagues what they find useful, rather than sorting out the deluge individually. An alternative might be YouTube-style ratings. EOs need a virtual repository to help users make sense of the volume of data.

Use of data

Accessibility is not simply about making data easy to find, but rather making it clear how to use the data. As one participant noted, "Knowing where the data is is important. Knowing what it *means* is key." Unfortunately, the way that data is described depends on the audience, and if the audience is unknown, it may need to be described differently. Over time, metadata needs to be supple enough for future use and for building backwards capabilities. Part of this process is recognizing the "value chain," whereby data is valuable to others down the line and making sure that the data can evolve to those needs. Even end products such as visualizations can be reused by interested parties. Providing this level of thoughtful curation, however, requires funding—and most research is not typically funded for broad use.

Once the data is found and understood, how do people validate that data, or any subsequent output, is trustworthy? A desirable goal is to put together "smooth exit" (validation) down the entirety of the line, from origin to citizen scientist. Libraries are currently helping scientist write metadata because of the difficulties in ensuring quality check and control, but even though guidelines and standards are be helpful, enforcement is necessary. Participants suggested that scientists want to know the underlying algorithms, and policy makers need to know that their arguments are appropriately supported. As a potential solution, it was suggested that there could be an Underwriters Laboratory[8] equivalent stamp of approval.

Tools

PIs are being challenged by NSF to consider how they will maintain infrastructure, including sustainable data management, over time. Individuals or research organizations that develop tools rarely have the resources to maintain or distribute it for general use. While federally-funded projects cannot charge fees for public data, soliciting corporate funding might be problematic, as profit motivated interests may make access too costly for smaller schools or museums. A possible solution to mitigating the cost of private

[5] NCAR Community Data Portal: http://cdp.ucar.edu/
[6] FriendFeed: http://friendfeed.com/
[7] Delicious: Social Bookmarking: http://delicious.com/
[8] Underwriters Laboratory, Inc.: http://www.ul.com/global/eng/pages/

sector tool development is through scaling different pricing models to the institutional size or type.

Use cases and engaging stakeholders

Stakeholders are the communities and their members that have a moral, legal, or ethical interest in a project. Stakeholders may include students, funding agencies, policy makers, citizen scientists, individuals, organizations, regional groups, multiple disciplines, but these parties are not necessarily apparent. Identifying stakeholders through the data centers that serve them is difficult because data centers are reluctant or unable to ask users for identification. Therefore, in the development of EOs, it is important to include sufficient funding for engaging, coordinating, and collaborating with stakeholders. One way that some initiatives are allowing multiple groups to discuss requirements is through use cases.

FEON as research council

Participants also discussed how FEON could address some data accessibility and distribution issues and challenges. One idea was to establish FEON as a research council, much like industries who band together to do research. It would be a resource for advice or "for hire" to conduct research as well as a data repository to help researchers and students find reliable data. The disincentives for using FEON as a research hub include financing, inherent lack of innovation and creativity for the organizational structure, and failure to understand the value chain allowed by sharing data.

Appendix A: List of Participants with Affiliation

First Name	Last Name	Institution	Department
Ilkay	Altintas	SDSC/UCSD	
David	Arctur	OGC Interoperability Institute	
Steve	Aulenbach	NEON	
Peter	Backlund	NCAR	
Karen	Baker	UCSD	SIO
Kathleen	Baker	Western Michigan University	Geography
Roger	Barry	Univ. of Colorado	CIRES
Mark	Bradford	NCAR	EOL
James	Brunt	LTER Network Office	Biology MSC03 2020
Lawrence	Buja	NCAR	
Jeffrey	Campbell	UMBC	Center for Urban Environmental Research and Education
Cyndy	Chandler	Woods Hole Oceanographic Institution	Marine Chemistry and Geochemistry
Sandeep	Chandra	UCSD, SDSC	
Alan	Chave	Woods Hole Oceanographic Institution	Applied Ocean Physics and Engineering
Patrick	Clemins	National Science Foundation	Division of Biological Infrastructure
Craig	Cochell	University of Oklahoma	
Gerald	Creager	Texas A&M University	AATLT
Theresa	Crimmins	National Phenology Network	
Mike	Daniels	NCAR/Earth Observing Laboratory	Computing, Data and Software Facility
Bob	Dattore	NCAR	CISL/DSS
Ewa	Deelman	University of Southern California	Computer Science
Ben	Domenico	UCAR	Unidata
Percy	Donaghay	University of Rhode Island	Graduate School of Oceanography
Robert	Downs	Columbia University	Center for International Earth Science Information Network
Peter	Fox	NCAR	

First Name	Last Name	Institution	Department
Janet	Fredericks	Woods Hole Oceanographic Institutions	AOPE
James	Gallagher	OPeNDAP	
Jeffrey	Goldman	UCLA	Center for Embedded Networked Sensing
Bette	Grauer	McPherson High School	Science Department
Sara	Graves	University of Alabama - Huntsville	Information Technology and Systems Center
John	Graybeal	Monterey Bay Aquarium Research Institute	Research and Development
Sandra	Henderson	UCAR	
Bill	Howe	Oregon Health & Science University	Center for Coastal Margin Observation and Prediction
Steven	Jackson	University of Michigan	School of Information
Melinda	Laituri	Colorado State University	FRWS
Venkat	Lakshmi	University of South Carolina	Geological Sciences
Katherine	Lawrence	University of Michigan	
Allen	Lee	Arizona State University	School of Human Evolution and Social Change
Kerstin	Lehnert	Columbia University	Lamont-Doherty Earth Observatory
Dan	Lubin	NSF	Office of Polar Programs
Deborah	McGuinness	Rensselaer Polytechnic Institute	Computer Science and Cognitive Science
Don	Middleton	NCAR	CISL/TDD
Barbara	Minsker	University of Illinois, Urbana-Champaign	Civil and Environmental Engineering/NCSA
James	Moore	NCAR/EOL	
James	Myers	NCSA/UIUC	
Daniel	Neafus	Denver Museum of Nature & Science	Planetarium
Brand	Niemann	US EPA	CIO
John	Orcutt	UCSD	Scripps Institution of Oceanography
Michael	Page	Emory University	Emory Libraries
Mark	Parsons	University of Colorado	National Snow and Ice Data Center
Marshall	Peterson	NEON Inc	

First Name	Last Name	Institution	Department
Paulo	Pinheiro da Silva	Univ. Texas at El Paso	Computer Science
Mohan	Ramamurthy	Unidata/UCAR	
Hampapuram	Ramapriyan	NASA GSFC	Earth Science Data and Information System Project
Brian	Rogan	Museum of Science	
Neil	Sarkar	Marine Biological Laboratory	MBLWHOI Library
David	Schell	Open Geospatial Consortium	
David	Schimel	NEON Inc and NCAR	
Lowell	Stott	University of Southern California	Earth Science
James	Syvitski	U. Colorado	CSDMS/INSTAAR
Alex	Talalayevsky	Ocean Leadership	OOI Cyberinfrastructure
Frank	Vernon	UCSD	SIO
Colin	Ware	University of New Hampshire	Center for Coastal and Ocean Mapping
Ronald	Weaver	National Snow and Ice Data Center	CIRES
Nancy	Wiegand	University of Wisconsin-Madison	Land Information and Computer Graphics Facility
Steven	Williams	NCAR	EOL
Chaowei	Yang	George Mason Univ.	CISC
Ilya	Zaslavsky	UCSD	San Diego Supercomputer Center
Jianting	Zhang	University of California at Davis	Dept. of Computer Science
Peisheng	Zhao	Center for Spatial Information Science and System	

Appendix B: Comments from Position Statements

Workshop participants were asked to submit a position statement prior to arrival. Position statements addressed the following three questions:

1. What are the exciting opportunities in cyberinfrastructure for environmental observatories?
2. What recent technical innovation do you see as most important?
3. Please describe the non-technical factors that you see as most important for successful cyberinfrastructure implementation.

The following summaries and excerpts convey the range of responses.

1. What are the exciting opportunities in cyberinfrastructure for environmental observatories?

- Capturing, sharing and using data in new ways, with new technologies by different users
- Advanced query/ search capabilities
- Model interoperation, data and models
- Real-time, and dynamic sensors
- Data and information to address society-critical problems
- Cultural and social aspects being addressed for interdisciplinary use and other audiences
- Cloud computing, web as the computer, source
- Doing 'things' on a large-scale, HPC, HPN, HPD
- Willingness to coordinate across EOs, federation
- Real applications of CI
- Meaningful semantic capabilities are starting to be available for science use

"The possibility of understanding our planet in ways never done before thanks to the possibility of large-scale integration of information, whether directly or indirectly related to measures from our environment." — Paulo Pinheiro da Silva

"Opportunities in cyberinfrastructure for environmental observatories are to provide web access to real time environmental data. This type of data when provided through an easily accessible portal and organized into user friendly analysis options can provide educators and their students with science applications that will stimulate student interest beyond that currently provided in the classroom. The data available from the environmental observatories can be used to compliment field visits or bring the field into the classroom." —Bette Grauer

2. What recent technical innovation do you see as most important?

- IT is cheaper for more, petascale, disk, net
- Near-real time, embedded systems
- Instrument/ detector innovations
- Improvements in realism of models

- Service-oriented architecture (SOA), web services, Representational State Transfer (REST)
- OGC protocols
- Semantic languages and tools, content and tools
- Workflow systems more widely used, some maturing
- Highly-capable digital repositories, content management systems
- Virtualization and anywhere-ness
- Data and tools sharing
- Federated identity and security
- Maturing and robustness of CI software
- Policy management of resources
- Better SPAM filters for e-mail
- Collaborative technologies, wikis, chat, AG, webcast
- Cool stuff - iPhone, GoogleEarth
- Web 2.0, social networking
- Citizen science, data generation/access
- Easier access to GIS

"High speed communications including high bandwidth not only permit rapid, global delivery of data, but large data products derived through the assimilation of near-real-time data. The availability of commodity computing and data clouds may represent a new opportunity for delivering, archiving, versioning, and modeling data." —John Orcutt

"Another important technological change has been portability which has allowed for data collection by far more individuals through citizen science initiatives. When students can contribute through such programs as GLOBE, they feel they understand the environment and why it needs to be stewarded. The innovations in portablility from laptops to handheld sensors has made this all possible." —Brian Rogan

3. Please describe the non-technical factors that you see as most important for successful cyberinfrastructure implementation.

- Training for users, staff, scientists
- Trickle-down of research CI to education and citizen access
- Existence, identification and engagement of stakeholders (in a sustained way)
- Understanding market forces; incentives and disincentives, buy-in factors, acceptance, trust, priorities, needs
- Seeking out end-users for new data products
- Adjusting to and creating new organizational structures, scaling across local, regional, national, global
- Enabling effective virtual collaborations (organizations), i.e. to work!
- Mechanisms for collaborative governance, coordination to share resources
- Integration of CI practitioners and science community
- Ability to understand interdisciplinary terminologies, communicate

- Connecting a diverse group of people with complementing skills (new, different), managing tensions and competing needs, mutual benefits
- Relations and synergies with peers in industry
- Alignment of disciplinary norms, routines, practices
- Overcoming NIH
- Willingness of developers and projects to adopt standards
- Development and acceptance of distribution standards for data, information
- Commit to full life-cycle project management for CI
- Reliable and easy-to-use computing environments
- Usability and reluctance to adopt new technology
- Funding allocations; software, maintenance, collaboration (long-term, esp. in the new teams)

"Recognition that the interdisciplinary nature of the problems we are researching means that we have to develop significant collaborations with new research communities that approach science very differently than our community does." — Lawrence Buja

"A quick list would include: trust (in data, in colleagues, in the integrity of data or knowledge produced in places and fields far from one's own); the alignment of disciplinary norms, routines, and practices (which turn out to show more variance across scientific fields than is typically imagined); the management of collaborative tensions (over resources, credit, patterns of exclusion, etc.); and appropriate mechanisms of collaborative governance (around, e.g., project decision-making, management of dissent or alternative visions, etc.)." —Steven Jackson

Appendix C: Agenda

May 5

12:30-12:45	Welcome, Logistics, Objectives (Peter Backlund, NCAR)
12:45-1:15	Overview of Key Issues Identified by Participants (Peter Fox, NCAR)
1:15-1:45	Report from recent Cyberinfrastructure for Environmental Observing Networks Meeting (Barbara Minsker, UIUC)
1:45-2:00	Charge to Breakouts
2:00-4:00	Breakout Sessions
	#1: Technology Challenges
	#2: Organizational Challenges
	#3: Standards Challenges
4:00-5:00	Plenary—Breakout Reports and Discussion

May 6

9:00-9:15	Review agenda
9:15-10:00	Semantic E-Science: Progress and Possibilities (Deborah McGuinness, Rensselaer Polytechnic Institute)
10:00-10:45	The View from NEON (Dave Schimel, NEON CEO)
10:45-11:00	Charge to Breakouts
11:00-12:00	Breakout Sessions
	#1: Instruments and Measurements
	#2: Data Processing, Assimilation, and Archiving
	#3: Information Distribution and Accessibility
12:00-1:00	Lunch
1:00 -3:00	Breakouts Continue
3:30-5:00	Plenary—Breakout Reports and Discussion

May 7

8:00-9:00	Continental Breakfast
9:00-9:15	Review Agenda

9:15-10:00 NASA's Earth Science Data Systems: A "Bit of History" and Observations (Hampapuram K. Ramapriyan, NASA)

10:00-10:45 Frontiers: Looking Back to Look Forward (Sara Graves, UA Huntsville)

11:00-12:00 Moderated Closing Discussion and Next Steps (NSF and Organizers)

Appendix D: Breakout Group Background Material

Technology Challenges

- Position statements, Question 1: What are the exciting opportunities in cyberinfrastructure for environmental observatories?
- Position statements, Question 2: What recent technical innovation do you see as most important?
- Adaptable cyberinfrastructure to respond to evolving science
- Need for a tailored product which provides trustworthy and understandable information
- Enable PIs to submit well described data
- Network interoperability
- Timeliness of data delivery
- Secure delivery of data
- Accessibility of data and tools by a diverse user base where not all are experts
- Support of multi-disciplinary research
- Verification of properly labeled data / Metadata standards
- Maintenance and reporting of provenance of data and derived products
- Exponential growth of data, scalability
- Establish a raw data center for the collection of data
- Create infrastructure for a single logical data storage location
- Storage and management of highly dimensional data
- Registration of data in four-dimensional space-time
- Authentication of system users to support multiple levels of access (read only, comment, edit, delete)
- Capture of data context including disturbance history and other historical aspects

Organizational Challenges

- Position statements, Q1: Willingness to coordinate across EOs, federation
- Position statements, Q2: Policy management of resources
- Position statements, Q3: [All answers]
  - Training for users, staff, scientists, ...
  - Trickle-down of research CI to education and citizen access
  - Existence, identification and engagement of stakeholders (in a sustained way)
  - Understanding market forces; incentives and disincentives, buy-in factors, acceptance, trust, priorities, needs
  - Seeking out end-users for new data products
  - Adjusting to and creating new organizational structures, scaling across local, regional, national, global
  - Enabling effective virtual collaborations (organizations), i.e. to work!
  - Mechanisms for collaborative governance, coordination to share resources
  - Integration of CI practitioners and science community
  - Ability to understand interdisciplinary terminologies, communicate
  - Connecting a diverse group of people with complementing skills (new, different), managing tensions and competing needs, mutual benefits

- Relations and synergies with peers in industry
    - Alignment of disciplinary norms, routines, practices
    - Overcoming NIH
    - Willingness of developers and projects to adopt standards
    - Development and acceptance of distribution standards for data, information
    - Commit to full life-cycle project management for CI
    - Reliable and easy-to-use computing environments
    - Usability and reluctance to adopt new technology
    - Funding allocations; software, maintenance, collaboration (long-term, esp. in the new teams)
- Governance structure for shared resources and collaboration
- Community buy-in to resources
- Incentives and culture change to make data sharing popular
- Rules vs. Incentives in terms of data sharing and community buy-in
- Ethical concerns about shared data
- Societal contributions, Community/Citizen science
- Promotion of sustained communication between EONs
- Keeping human expertise

Standards Challenges

- Overcoming NIH
- Willingness of developers and projects to adopt standards
- Development and acceptance of distribution standards for data, information
- Usability and reluctance to adopt new technology
- OGC protocols
- Semantic languages
- Standards body
- Sustainability / Maintenance
- Ontology for environmental and ecological variables
- Mission planning and optimization

Instruments and Measurements

- How do we capture data in new ways, with new technologies by different users?
- How to cope with real-time, and dynamic sensors (tasking)?
- What is commodity and what is not?
- How to accommodate and facilitate citizen science, citizen data generation?
- What standards are ready to use (can I buy them standards-enabled)?
- What metadata is captured and propagated?

Data Processing, Assimilation and Archiving

- Viability of virtualization, cloud computing, web as the computer and (potential) source of related data?
- How to utilize workflow systems, mining, agents and facilitate use with models?
- How to request and manage resources (yours or others) - taking advantage of high-performance computing, network, I/O, data streams AND low-performance !!

- How to adopt best practices for data preservation, archiving?
- What is the key lineage, provenance and derivation metadata?
- How to take advantage of digital repositories, content management systems?

Information Distribution and Accessibility

- How to determine existence and engagement of stakeholders (in a sustained way)
- How to understanding market forces; incentives and disincentives, buy-in factors, acceptance, trust, priorities, needs, social and cognitive factors?
- How to seek out end-users for new data products?
- How to enable interdisciplinary use esp. for other audiences: semantics, viz. tools, simple ones
- What are the advanced query/ search capabilities?
- How to provide Data and information as a service?