Improving Observing Network Coordination: A Cyberinformatics Forum

Workshop Report

May 17–19, 2010

National Center for Atmospheric Research

Boulder, Colorado

Report prepared by

Peter Backlund, National Center for Atmospheric Research

Peter Fox, Rensselaer Polytechnic Institute

Greg Guibert, National Center for Atmospheric Research

Introduction:

In May 2010, the National Center for Atmospheric Research (NCAR) hosted a cyberinformatics forum in Boulder, CO on Improving Observing Network Coordination. The forum was the third and final event organized under a National Science Foundation (NSF) grant to provide guidance and assistance in the design, management, and coordination of environmental observing networks across the nation. Building on discussions, recommendations, and inputs from the previous two forums in 2006 and 2008, as well as nearly a decade of effort by the cyberinformatics community to address the technical, management, and development challenges of integrating and coordinating observing network efforts, the forum resulted in a number of suggestions and recommendations to address barriers to coordination as well as facilitate partnership across observing networks. The goal of the forum, however, was not to generate a consensus for action or to develop an actionable set of recommendations, and instead was an open dialogue among participants from a wide range of scientific and technical disciplines, representing both large and small observing networks funded by NSF, other federal agencies, and third-party or private sector initiatives.

This report details the plenary and breakout session discussions and outlines the overall conclusions from the forum. Additional background information, including the final agenda and list of participants, is located in the appendices. The forum was designed specifically to generate conversations around a focusing set of questions with the goal of providing a common framework for addressing the complex interlinkages between the technical, data, and management issues associated with coordination across diverse networks. Throughout the course of two days, the results of breakout session discussions were fed back to the entire group and served as inputs for the further rounds of breakout discussions and refinements to the overarching theme of improving network coordination.

The forum built off of several years of discussions and meetings within the CI community on overcoming the challenges of increasing integration and cooperation between efforts. Over the course of these engagements, different mechanisms have been explored, most notably the creation of a Federated Environmental Observatory Network (FEON). FEON was the topic of a separate NSF sponsored workshop in February 2008 and was prominently featured in the May NCAR 2008 Cyberinformatics Forum. Questions about funding and support resources, specific implementation mechanisms, and the top-down role in the development and diffusion of standards have led to the community losing some enthusiasm for this structured solution to the cooperation issues. While the FEON concept continues to influence conversations about the type, scope, and scale of the collaborative mechanisms being explored, FEON itself as a mechanism was not discussed in depth at the 2010 Forum.

Content, Methodology, and Participants:

Participation in the forum was open to all members of the community and was sought broadly through an open call via community listservs, email solicitations, and online forums over several months. When registering, participants were asked to provide some initial thoughts on four questions (listed below) designed to elicit early thinking on the opportunities, challenges, barriers, and innovations associated with the coordination of observing networks:

> **Question 1:** Do you think the NSF-funded observatories need a dedicated coordination mechanism for CI and other issues? Or, could they achieve sufficient coordination through participation in existing multi-agency coordination efforts?
>
> **Question 2:** What near-term opportunities are there for joint or shared CI across NSF-funded environmental observatories?

Question 3: Are there new technical and/or programmatic developments that have potential to ease coordination and collaboration among environmental observing projects and user access to environmental observations?

Question 4: Do you have specific suggestions for improving the development and use of CI in NSF-funded environmental observing projects?

The forum included plenary and breakout sessions. The breakout discussion sessions linked to the four focus questions and associated participant responses, and further supported by plenary reporting and feedback sessions, as well as several formal presentations on specific ongoing community coordination efforts. There were four formal plenary presentations:

- Dr. Meredith Lane from the National Biological Information Infrastructure (NBII) provided a detailed look at the NBII initiative.
- Dr. Mark Williams from the University of Colorado described the new Boulder Creek Critical Zone Observatory (CZO) and the overall CZO project.
- Dr. Ted Haberman from the National Oceanic and Atmospheric Administration discussed the ongoing development and application of ISO data standards.
- Dr. Carol Myer provided an overview of the ESIP Federation, a consortium of 115 Earth and environmental data centers formed in 1998 to increase the use and impact of Earth science data and information.

Full presentations are available at www.cyberobservatories.net

The breakout sessions were loosely moderated to facilitate conversation but ultimately the discussions were organic, wide-ranging, and driven by the priorities and perspectives of the participants.

Participants included staff from NSF-funded environmental observing projects (both "domain" scientists and CI practitioners), staff from environmental observing projects funded by other agencies, computer and computational scientists, software engineers, and academics and students.

Summary of Main Conclusions:

The Cyberinformatics Forum sought to address the challenges, opportunities, and barriers to one of the most pressing and persistent issues facing the earth observing community: cross-institution collaboration. Unlike previous forums that focused on the development of a large-scale formal collaboration structure, such as the FEON concept, the discussions at the 2010 Forum identified some specific and actionable mechanisms that, when knitted together, would have direct impact on the level of interaction between observatories. Following is a brief description of ideas that generated broad support and enthusiasm and/or widely shared concerns:

Overall, there were mixed feelings about the need for a dedicated mechanism to coordinate the CI component of NSF-supported observing systems, although there was a strong consensus on the value of and need for coordination. A suspicion of formal, top-down coordination was countered with a strong preference for a more informal bottom-up approach. Coordination of observing CI was seen as a need and an opportunity, but that many other aspects of observing systems need to be coordinated as well and CI should not be seen as an separate, stand-alone issue.

Coordination among small-scale and large-scale observing programs was also identified as a pressing need. Significant attention is appropriately focused on large spatial scale, large budget observing projects, but many science challenges require data and observation on much smaller

scales, integration of data from multiple small-scale observing systems, or integration of data from small and large scale systems. Neither large nor small systems have a monopoly on innovative tools and methods.

Although the NSF-supported projects share certain affinities and needs, there are many other observing projects supported by other agencies, and NSF projects have lessons to share, as well as learn, within this broader context. The Earth Science Information Partnership (ESIP), for example, that grew out of NASA but now includes NASA, NOAA, and DOE-supported projects was discussed at the meeting and seen as a valuable coordination mechanism that cuts across agency projects. At the same time though, it is also seen as nearing its capacity and therefore vulnerable to being overloaded and made less effective. Because many CI experts are active across a range of NSF, NASA, NOAA, DOE, and USGS-supported projects, it was felt that a coordination mechanism should be designed to nurture and bring together a larger community rather than segmenting the science community along agency lines.

Despite the promise of increased coordination, it was acknowledged that there were limits to the potential extent of collaboration across platforms. In the end, a range of coordination mechanisms will be needed within and across agency communities, communities of scientific and technical practice, and large and small projects and programs. Much of the coordination will grow organically in response to specific needs and challenges. The support of NSF and other agencies should therefore focus on nurturing such interactions and avoid, to the extent possible, the creation of groups or mechanisms that are not grounded in specific needs identified by the scientific community.

Most participants felt that a periodic (yearly or every other year) large community meeting to allow CI research and practitioners across disciplines to share research, best practices, and innovations could be valuable. Such a gathering could be part of the answer to the need to "coordinate the coordination mechanisms" as well as the observing and CI projects. Most participants applauded the increased emphasis on observational CI, and CI in general, at AGU and other large-scale gatherings, but most felt that this was not adequate for focused interaction. However, the enthusiasm for a focused gathering was tempered by a strong feeling that the meeting calendar is already oversaturated. If a single larger meeting could reduce the overall number of meetings by replacing some small-scale meetings it would be more desirable. Because it was felt that promising virtual social networks and technologies were dependent on initial levels of person-to-person interaction, a community scale meeting could serve an important introductory purpose, as well as sustaining and deepening existing interactions. If combined with the potential to access new, small grants to operationalize the ideas and innovations that flow from these interactions, there was a feeling that concrete follow-up and momentum would alter the traditional, relatively passive dynamic of large conferences and workshops. Instead, a community scale conference could be seen as a true catalyst for innovation and scientific progress.

There was strong support among participants for creation of a "CI Synthesis Center" that would promote innovation, sharing, and adaptation of CI tools and methods to specific domain science challenges, including observing systems. Such a center would act in the spaces between observatories and CI projects, and be somewhat analogous to existing scientific synthesis centers. It would serve as a clearinghouse for best practices, solutions and innovations, perhaps with an "open code repository" to support sharing and repurposing, and as a community gathering space where groups could convene for specific integration projects or conduct targeted "hack-a-thons" for well-defined CI/domain problems. The center would be characterized by loose, meso-scale forms of interaction around CI development and would be far less structured than formal/grant-based mechanisms.

There was also strong support for a funding mechanism to provide small, supplemental additions to existing grants that would be directed at specific collaborative enterprises between two or more CI initiatives. It was felt that the supplements should be for creating linkages that were very specific and tangible so that could not be absorbed into existing funding streams with little return on new investment. If combined with the creation of a CI Synthesis Center, the small grants could be disbursed via the Center, thereby helping alleviate concern that the funding would be absorbed into an observatory's core funding. Additionally, the financing of targeted interaction between two observatories with concrete partnership objectives was felt to have significantly higher value than large non-specific community networking meetings or even smaller thematic forums with a diverse groups of invitees.

Participants were very interested in and concerned about the new NSF data management policy. If the requirement for data management plans is coupled with funding for implementation and consistent guidance from program managers, it could greatly facilitate the coordination of observing systems. If grantees are required to seek and adopt a data management system, they will, out of cost necessity and time constraints, look to satisfy these demands as efficiently as possible. Ultimately, this new pressure will result in some management standards becoming more prominent and dominant than others, essentially creating a peer-initiated weeding process of standards and approaches. In contrast, implementing the policy as an unfunded mandate is likely to result in uneven application across programs, resulting in more confusion and could, in fact, damage nascent community-based coordination efforts.

Discussion of Key Questions:

The centerpieces of the forum were several rounds of small group breakout sessions whose goal was to explore in detail the issues raised by the questions and the formal plenary presentations. The summaries of issues, ideas, and concerns below draw on the breakout discussions and the responses submitted prior to the forum. They do not necessarily represent consensus opinions and, in fact, many of the discussions summarized may conflict with one another.

Question 1: *Do you think the NSF-funded observatories need a dedicated coordination mechanism for CI and other issues? Or, could they achieve sufficient coordination through participation in existing multi-agency coordination efforts?*

Central to the responses to Question 1 was the definition, scope, and scale of the "coordination". To the extent that coordination would be represented by an increase in inter-observatory communication, there was general enthusiasm for increased mechanisms for voluntary engagements. However, when coordination was defined more structurally as a series of mandates, standardizations, and new institutional entities, there was concern that innovation and organic growth would be stifled. In most cases, bottom-up options were favored over top-down mechanisms, with an emphasis on advisory or guidance structures rather than those that would be proscriptive.

Collaboration generally could be seen occurring across five different sectors: science, data, tools, experience, and IT infrastructure. In each case, there were historical inhibitory factors that reduced the ability to share or collaborate across observatories. For example, data standards have failed to evolve naturally because there is variable maturity between efforts that result in uneven utility across domains and tools. As a further consequence, the time, money, and political resources needed to create standards through a systemized approach has proven intractable.

While increased communication and interaction was often cited as a benefit and need of a

coordination effort, there was a general feeling that, whether a new mechanism was adopted or existing efforts were modified to address coordination needs, care should be taken not to proliferate additional meetings, conferences, and forums that would place additional time demands on community members. There was a strong feeling that both too much and too little time spent coordinating is a problem. Although the increased exposure of cyberinformatics at big science meetings such as AGU was seen as advantageous, the trend is not sufficient to cover the breath and depth of topics needed by the community. At the same time the proliferation of smaller meetings aimed at CI management issues was seen as overwhelming and ineffective. One suggestion, although not embraced by all participants, was the potential for one large periodic CI community meeting where researchers and technicians were able to present detailed descriptions of their actual projects and results directly to their peers.

There were a number of additional benefits associated with a coordinated mechanism that were identified. Some participants felt that a coordinated mechanism would result in cost efficiencies and substantial reduction in the duplication of effort in the creation and maintenance of cyberobservatories and the data they generate. Another benefit is the potential to increase the user base of cyberobservatories and their data products through the consistency of data presentation and added dissemination capacity. It would also potentially be easier to have a central user group body to steer funding towards, leveraging larger group capacities rather than stand alone projects.

Some expressed hope that communication between CI efforts would be naturally enhanced through the better use of existing efforts (e.g. ESIP, DataONE and other DataNet projects, OGC, LTER Network Office, TDWG, National Middleware Initiative, ITIS, GEOSS). Rather than generating large, expensive new mechanisms, an assessment of the existing capacities and gaps could be used as a blueprint for the creation of some new, small, targeted efforts to fill the spaces between programs and thereby bring domains and disciplines closer together. For example, a series of small supplemental grants could be deployed specifically to create joint development opportunities and to innovate around the edges of existing programs with the goal of pulling in other CI initiatives where they might not otherwise be engaged. There was also concern that large coordination efforts had been attempted in the past, with limited success, and that those lessons had not necessarily been well learned.

Question 2: *What near-term opportunities are there for joint or shared CI across NSF-funded environmental observatories?*

Many of the participants felt that there were substantial near-term opportunities for joint or shared CI, some of which were already being taken advantage of while others could be considered ‘low hanging fruit’ and not yet developed. However, in many sessions, the discussion began with an analysis of the barriers to sharing CI across observatories.

There was a general acknowledgement that differing project goals, needs, and schedules often created mismatched objectives and project timing when considering sharing CI across observatories. Additionally, grant sponsors and reviewers were uncomfortable supporting shared infrastructure or endorsing projects that would rely on outside items for critical path activities. For technicians and researchers, time constraints and a lack of clear reward structure also inhibits cross-observatory collaboration.

There was broad agreement that there is increasing recognition at NSF of the importance of CI/Informatics for observing programs and that program managers are already pushing observing programs together through a variety of small or incremental actions. At a broader level, OCI is also working more across directorates to provide increasing opportunities.

When considering what resources are already being shared across observatories, it was mentioned that people themselves represent one of the most valuable CI assets. Researchers and technicians are often co-associated with different efforts, resulting in a valuable, but difficult to measure, sharing of information and attitudes across networks. This is contributing to the re-purposing of tools and methods that virtually all workshop participants see as critical for efficient and rapid progress in this field. The development of specific social networks, both virtual and face-to-face, between technicians and between managers could further facilitate collaboration, communication, and sharing.

It was also suggested that the observatories could look to non-science community activities and data systems to see what they are using and how they generate new ideas. In many instances there are tools, services, and formats that are defacto standards because of their widely accepted use or market penetration and should be formally acknowledged and adopted rather than perpetuating a debate about new standards.

With the increasing recognition of the value of data and data providence there is a potential role for Earth Observatories to serve as providers of archive services. Many participants felt that the rapid ongoing development of high-level metadata standards, controlled vocabularies, and convergence of metadata languages held considerable promise for easing the challenge of discovery, integration and analysis of data sets from multiple sources.

Question 3: *Are there new technical and/or programmatic developments that have potential to ease coordination and collaboration among environmental observing projects and user access to environmental observations?*

There was a general acknowledgement that advances on several technical and programmatic fronts have the potential to significantly ease coordination and collaboration among environmental observing projects. Although often mutually supporting, this section is divided roughly into the new technical and new programmatic developments that were identified and discussed. In several discussions, it was noted that there are a number of community driven to share CI already exist, including NCAR, TDWG, Iris, ESIP, and OGC. The community-driven nature of these existing efforts typically augurs well for their long-term sustainability, design, and buy-in versus the creation and imposition of standards from a centralized entity.

New Technical Developments:

Many of the new technical developments that were seen to have potential to ease coordination involved social networking and associated sharing technologies. Twitter, Skype, net relay chat, listservs, and group sharing and reviewing of documents were among the technologies identified that could be seen facilitating the easy and inexpensive flow of information between individuals and groups. It was noted, however, that internet and virtual networks and tools alone were insufficient and typically relied on some initial face-to-face interaction to initiate and sustain contact and collaboration.

A number of specific technical advances were also identified. For instance, some participants felt that cloud computing could provide a solid foundation for increased flexibility, interconnectivity, interoperability, and evolution. But participants also cautioned that this could expose projects to future cost increases that could prove difficult to manage.

Advances in ISO standards and the Open Provenance Model (OPM) were discussed for their low development overhead and ease of use/maintenance over time. The movement towards open-source technology development, with broadly accessible code-repositories, bug-tracking systems, and public forum, for example, was seen by some as enabling and simplifying

coordination and discussion among geographically remote participants.

The rapid evolution of new mobile technologies, facilitated by high-bandwidth networks like Internet2, have provided a set of basic and powerful tools that can be combined and designed to create elegant, robust interfaces to large data collections. Geospatial support has typically required customized implementation and often suffered from performance issues that prohibited simple data access, but the broad use of geolocation features on cellular devices has necessitated new solutions that will benefit the entire observational CI community. Advances in database technologies that support geospatial / geodetic data also allow observatories to very efficiently catalog and search through a large amount of data.

New Programmatic Developments:

Several participants mentioned the new NSF data management policy, if backed by funding for implementation and consistent guidance from program managers, could greatly facilitate the coordination of observing systems. If grantees are required to seek and adopt a data management system, they will, out of cost necessity and time constraints, look to satisfy these demands as efficiently as possible. Ultimately, this new pressure will result in some management standards becoming more prominent and dominant than others, essentially creating a peer-initiated weeding process of standards that are less useful for boarder applications across domains.

In contrast, other participants were concerned the currently unfunded policy would result in uneven application across programs, resulting in more confusion and could, in fact, damage nascent community-based coordination efforts. There was lengthy discussion about how the CI community could support and positively influence the new policy and whether the new policy should use incentives or enforcements as a guiding philosophy. Leveraging the experience of the community to advise NSF program managers on raising metadata standards and to educate them on how to require PIs to contact/understand what already exists was seen as a positive, proactive role to assist with the implementation of the new policy.

Question 4: *Do you have specific suggestions for improving the development and use of CI in NSF-funded environmental observing projects?*

Participants offered and discussed a number of specific ideas over the course of the forum. One idea, however, generated considerable discussion in several breakout groups and in plenary sessions. Although there was broad consensus that a centralized, formal CI facility charged with establishing standards and protocols to be pushed into the community was not desirable, a nearly-virtual center whose goal was to synthesize research and practice could be valuable. As envisioned, a CI Synthesis Center would be analogous to existing scientific synthesis centers and serve as a clearinghouse for best practices, solutions, and innovations but not be a physical location, entity, or institution. This could perhaps include an "open code repository" to support sharing and repurposing. Through the center groups could quickly convene for specific integration projects or conduct targeted "hack-a-thons" for well-defined CI problems. The center would be characterized by loose, meso-scale forms of interaction around CI development and would be far less structured than formal/grant-based mechanisms. It was noted that Global Biodiversity Information Facility (GBIF) and the National Biological Information Infrastructure (NBII) are doing some aspects of this now but that it has not been formalized or fully realized. Several of the other specific suggestions noted throughout the forum were seen as easy to integrate into a CI Synthesis Center. For example, small grants to fill spaces between efforts could be undertaken under the auspices of a synthesis effort.

A synthesis center would also benefit from low overheads and by being community driven without current observatories feeling as if they were working under the auspices of another

program's structures and branding, earlier identified as a significant barrier to collaboration. There was also a sense that one or two full time individuals whose job was dedicated to linking and networking the observatories to achieve specific collaborative goals could have a substantial impact on coordination, whether affiliated with a synthesis center or a separate mechanism. Similarly, a small amount of technical capacity to create resource or tool wikis or a YELP! for CI software and tools would also be a moderate investment in time and money and could return substantial dividends in providing a specific community-wide tool with immediate utility and could be seen as a web-based component of a synthesizing activity.

Appendix A – Final Agenda

Monday, May 17		**Speaker**
12:00 – 1:00	Opening Lunch	
1:00 – 1:15	Introduction, Review of Meeting Objectives	Peter Backlund, NCAR
1:15 – 2:15	Program Highlight: National Biological Information Infrastructure	Meredith Lane, NBII
2:15 - 3:00	Highlights of Participant Submissions	Peter Fox, RPI
3:00 – 3:15	Instructions to Breakouts	Peter Backlund, NCAR
3:15 – 3:30	Move to Breakouts	
3:30 – 5:30	Breakout sessions on Usability, Productivity, Linkages	
5:30 – 6:30	Break	
6:30 – 8:30	Welcome Reception	

Tuesday, May 18		
8:30 – 9:00	Breakfast	
9:00 – 10:30	Report and Discussion of Breakout Results	
10:30 – 11:00	Break	
11:00 – 12:00	Program Highlight: Critical Zone Observatory Project	Mark Williams, CU
12:00 – 1:00	Lunch	
1:00 - 1:45	ISO 191115-2 for Observations	Ted Habermann, NOAA
1:45 - 2:00	Instructions to Breakouts	Peter Fox, RPI
2:00 – 3:00	Breakout Sessions on Coordination and Collaboration	
3:00 - 3:30	Break	
3:30 - 5:00	Breakout Sessions continue	
5:00 - 6:00	Break	
6:00 – 8:00	Reception: Earth Science Information Partners	Carol Meyer, ESPI

Wednesday, May 19

8:30 – 9:00	Breakfast
9:00 – 10:30	Report and Discussion of Breakout Results
10:30 – 11:00	Break
11:00 – 12:00	Summary, Next Steps

Appendix B – Final Participant List with Affiliations

First Name	*Last Name*	*Affiliation*
Luis Marcelo	Achite	Inter-American Institute for Global Change Research (IAI)
Dan	Andresen	Kansas State University
Steve	Aulenbach	NEON
Peter	Backlund	NCAR - ISP
Kathleen	Baker	Western Michigan University
Archer	Batcheller	U. of Michigan
Willaim	Bergen	OOI/Raytheon
Luis	Bermudez	Southeastern Universities Research Association
Steven	Berukoff	NEON
Mark	Bradford	NCAR
Benjamin	Branch	Elizabeth City State University
Eric	Bridger	Gulf of Maine Research Institute
James	Brunt	Long Term Ecological Research Network
Lawrence	Buja	National Center for Atmospheric Research
Randy	Butler	University of Illinois
Bonnie C.	Carroll	National Biological Information Infrastructure
Yoori	Choi	CUAHSI
Mike	Daniels	NCAR/EOL
Steve	DeLong	University of Arizona
Sheldon	Drobot	NCAR
Peter	Fox	RPI & Tetherless World Constellation
James	Frew	University of California, Santa Barbara
Christian	Fritz	University of Southern California
Felimon	Gayanilo	University of Miami

Arthur	Gaylord	Woods Hole Oceanographic Institution
Henry	Gholz	National Science Foundation
Yolanda	Gil	University of Southern California
Greg	Guibert	NCAR – Integrated Science Program
Ted	Habermann	NOAA/NESDIS/NGDC
Tony	Hays	NEON
Matthew	Howard	Texas A&M University
Kolby	Jardine	University of Arizona-Biosphere 2
Scott	Jensen	Indiana University
Matthew	Jones	University of California, Santa Barbara
Thomas	Kampe	NEON, Inc.
Eric	Kihn	NOAA
Jack	Kleinert	Raytheon Corporation
Meredith	Lane	National Biological Information Infrastructure
Kerstin	Lehnert	Columbia University
Christopher	Lenhardt	Oak Ridge National Laboratory
Doug	Lindholm	Laboratory for Atmospheric and Space Physics (LASP)
Andrea	McCurdy	Ocean Leadership
Michael	McMahon	University of Nevada, Reno
Carol	Meyer	ESIP Federation/Foundation for Earth Science
Don	Middleton	NCAR
Jim	Myers	U. Illinois
David	Neufeld	CIRES
Michael	Palecki	National Climatic Data Center
Carol	Park	NCAR
Mark	Parsons	University of Colorado

Michelle	Rangel	NCAR - ISP
Rob	Raskin	Jet Propulsion Laboratory
Rob	Redmon	National Oceanic Atmospheric Administration
Allan	Reynolds	NEON
Paula	Robinson	NCAR - ISP
Alyssa	Rosemartin	USA National Phenology Network
Mark	Schildhauer	National Center for Ecological Analysis and Synthesis
Trina	Shartsis	University of Arizona
David	Tarboton	Utah State University
Karan	Vahi	USC Information Sciences Institute
Frank	Vernon	UCSD
Ronald	Weaver	NSIDC-CIRES; University of Colorado
Brian	Wee	NEON, Inc.
Jake	Weltzin	USA National Phenology Network
Patrick	West	Rensselaer Polytechnic Institute
Scott	Wiant	NEON
Olga	Wilhelmi	NCAR
Mark	Williams	University of Colorado, Boulder
Anne	Wilson	CU Boulder
Ilya	Zaslavsky	UCSD
Stephan	Zednik	Rensselaer Polytechnic Institute
Mikhail	Zhizhin	NGDC NOAA