Insights and Precepts from a Workshop on Creating a Software Institute for Environmental Observatories

Workshop PI: Stan Ahalt[1]
Workshop Co-PI's: Barbara Minsker[2], Ray Idaszak[1] (editor)

Workshop Organizing Committee
Stan Ahalt[1], Barbara Minsker[2], Ray Idaszak[1], Dave Tarboton[3], Michael Tiemann[4], James Brunt[5], Robert Tawa[6], Peter Backlund[7], Mark Williams[8], and Mark Parsons[9]

1.0 Overview

In October 2010, RENCI[1] and NCSA[2] hosted an NSF-funded workshop[10] that explored creation of a Software Institute for Environmental Observatories. Representatives from nine NSF-funded environmental observatories, consortia, and observatory-related cyberinfrastructure projects attended: AON[11], CUAHSI CHyMP[12], CUAHSI HIS[13], CZOs[14], HydroNexrad[15], LTER[16], NEON[6], OOI[17], and WSC[18]. We recognize that strong alignments exist between our workshop and the NSF Earth Cube Initiative, both with the themes we've selected and the communities we are targeting. Several key insights and precepts derived from the workshop will likely be valuable to the Earth Cube community and for that reason, we are presenting an abbreviated version of the workshop report within this white paper. No single person authored this white paper; rather, we recognize the Principal Investigators and Organizing Committee members listed above, the workshop attendees listed in Appendix A, and the NSF attendees for their combined contributions to the original workshop report that served as the foundation of the content presented in this paper. The original workshop report in its entirety can be found here: http://www.acsu.buffalo.edu/~abani/RENCI_NCSA_NSF_1049273_workshop_report.pdf .

The workshop provided an open forum for the attendees to explore realization of the NSF CF21 vision and how to yield robust, sustainable products and methods that respond to the growing cyberinfrastructure

[1] Renaissance Computing Institute; http://www.renci.org/
[2] National Center for Supercomputing Applications; http://www.ncsa.illinois.edu/
[3] Utah Water Research Laboratory Civil and Environmental Engineering Department, Utah State University
[4] Open Source Affairs, Red Hat
[5] Long Term Ecological Research Network Office, University of New Mexico
[6] National Ecological Observatory Network (NEON); http://www.neoninc.org/
[7] National Center for Atmospheric Research (NCAR); http://ncar.ucar.edu/
[8] Institute of Arctic and Alpine Research, University of Colorado at Boulder
[9] National Snow and Ice Data Center, University of Colorado at Boulder
[10] http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=1049273
[11] Arctic Observing Network; http://www.aoncadis.org/
[12] Community Hydrologic Modeling Platform; http://www.cuahsi.org/chymp.html
[13] CUAHSI Hydrologic Information System; http://his.cuahsi.org/
[14] Critical Zone Observatories; http://criticalzone.org/index.html
[15] Hydro Next-generation Radar providing radar-rainfall data for use in hydrology, hydrometeorology, and water resources; http://hydro-nexrad.net/
[16] Long-Term Ecological Research; http://www.lternet.edu/
[17] Ocean Observing Initiative; http://www.oceanleadership.org/
[18] Water Sustainability and Climate Program; http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503452

needs of the environmental observatory communities. It addressed community architectures driving effective governance models for coordination of sustainable community practices, including integration of community workflows, interoperability of resources, and virtual organization principles that are expected to drive new science at the interstices of these communities. The workshop provided the opportunity for open discussion among leaders at each of the environmental observatories, cyberinfrastructure experts, and industry experts on the topic of what constitutes a successful Software Institute for environmental observatories. In addition to representatives from environmental observatories, vendor representatives were Red Hat, Microsoft Research, IBM, Environmental Systems Research Institute (ESRI), and Danish Hydraulic Institute (DHI). Other organizations represented included the National Center for Atmospheric Research (NCAR), Unidata, California Institute for Telecommunications and Information Technology (Calit2), San Diego Supercomputer Center (SDSC), National Snow and Ice Data Center (NSIDC), Stroud Water Research Center, and several universities where cyberinfrastructure research and development for environmental science is conducted. Reed Beaman, Alan Blatecky, Christy Geraci, Bruce Hamilton, Nancy Huntley, Cliff Jacobs (who delivered the NSF plenary presentation), Peter McCartney, Manish Parashar and Kevin Thompson attended from the NSF. In total there were 49 people who attended in person including nine from the NSF, and an additional four remote sites participated in the workshop via videoconferencing.

2.0 Workshop Key Organizational Precepts, Conclusions, and Recommendations

This section discusses the key organizational precepts, conclusions, and recommendations from the workshop. Sections 2.1 to 2.6 correspond to six general areas described in a NSF Dear Colleague Letter regarding Scientific Software Innovation Institutes[19]. We have rephrased the wording of these six areas herein to relate specifically to the environmental observatory communities. The subject matter for each section was derived from the entire workshop, including plenary and panel presentations, breakout sessions, and open discussions. ***Explicit recommendations to inform the NSF are indicated in BOLD ITALICS.***

2.1 How will significant challenges faced by the environmental observatory community benefit from an institute in terms of scientific innovation as well as productivity?

- **Software Ecosystem for New Science –** A primary tenet of the workshop was that data, software for data, and application software are inextricably linked. Many common issues faced by NSF-sponsored observatories in offering data and software services in support of grand challenge science would benefit from common software research and development coordinated by the Institute. The Institute for environmental observatories would provide a critically necessary software ecosystem that enhances and accelerates new data-driven science within communities while enabling new science between communities. The Institute would facilitate sharing of data elements via associated interoperability of software for data. There was an assertion that significant new grand challenge science and innovation will happen at the interfaces or interstices between communities. The need for an environmental observatories Institute was emphasized as a vital necessity to marshal the software assets enabled by NSF funding and that it be sustained for future scientific explorations not otherwise possible.

- **Synthesis –** An environmental observatories Institute would serve as a synthesis center enabling multi-disciplinary collaboration on critical grand challenge science such as mitigation and adaptation to climate change, global hypoxia, and ecosystem and water sustainability. In addition, support for cross-disciplinary working groups to collaborate in manipulating heterogeneous data

[19] http://www.nsf.gov/pubs/2010/nsf10050/nsf10050.jsp

from multiple sources into new scientific methods and knowledge is critical for progress on these challenges. That is, while *data is the idiom of environmental grand challenge science, software is its modality*. ***It is critically important that the NSF recognize the importance of the Institute serving the role of a true cross-cutting multi-functional institute in expanding synthesis within and across communities through data and software interoperability, support for synthesis working groups, and sharing of best practices.***

- **Software for Data –** Each of the environmental observatories recognized inherent data challenges within their respective communities. The potential benefits of the Institute to the environmental observatory communities were largely oriented around data services, including improved data handling and standards for interoperability, ease of use, and collaboration through data. A particular need was discussed for an infrastructure to enable community members to easily create and share custom data products that integrate diverse data types at multiple scales and are more easily used for analysis and modeling, and estimating the state of environmental systems across space and time (including error bounds). A recent memorandum from the Executive Office of the President regarding U.S. Science and Technology Priorities for the FY2012 Budget reinforces the priority of sustainability and interoperability of data:

  > "Agencies...should develop and sustain their datasets to better document Federal science, technology, and innovation investments and to make these data open to the public in accessible, useful formats. Agencies should develop and regularly update their sharing policies for research performers and create incentives for sharing data publicly in interoperable formats to ensure maximum value..."[20]

  There was discussion on the need for the Institute to provide software as infrastructure to facilitate these requirements both within and between communities, and on the importance of clearly differentiating the role of the Institute versus the roles of other NSF data-centric programs such as DataNet. ***An environmental observatories Institute should connect to DataNets and other data-generating initiatives and provide and sustain a software ecosystem that promotes an interoperable suite of customizable data products that are closely coupled with corresponding analytic and synthesis activities.***

- **Productivity and the Transformation of Science –** Productivity improvements and the transformation of science would result from more data of higher quality and greater usability from more sources. Currently, finding and transforming available environmental data for use in scientific studies and modeling can require months of effort for every research project, while readily available and reliable data products from an environmental observatories Institute could reduce this time to minutes. The need for increased productivity and enabling research in transformative science is a definite prerogative of the Institute.

2.2 What are potential key attributes of the Institute that would benefit the environmental observatory communities? What are the appropriate Institute organizational, personnel and management structures, as well as operational processes for this community?

[20] Orszag, Peter R., Holdren, John P. Science and Technology Priorities for the FY 2012 Budget. Memorandum For the Heads of Executive Departments and Agencies, Executive Office of the President, Washington D.C., July 21, 2010, pg. 2. Retrieved October 14, 2010 from http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-30.pdf

- **Interoperability –** Interoperability was especially significant in that some workshop attendees described interoperability *as the grand challenge* including how to bridge across the various communities. There was considerable discussion, but not consensus, on the approach. Some proposed a full-blown institute providing SaaS (Software as a Service) relying on resource providers where environmental observatory services could run primarily on academic or commercial resources not owned by the community, and where the Institute would coordinate data management. Others wanted a "lighter" solution with a focus on resource coupling, resource pooling and standards. The recommendation that distilled out of this discussion is that there are a variety of possible approaches to implementing the Institute, whether for environmental observatories or another focus.

- **Software Sustainability –** ***A clearly identified issue with sustainable software is that support for grant-developed software is typically nonexistent at the end of a grant. This is a key issue for the Institute to address. The Institute needs a clear method for encouraging and supporting the transition of successful software into sustainable operational infrastructure, and if possible, additionally provide substantive examples demonstrating that the proposed method has been successfully utilized in the past in achieving sustainability.***

- **Support for Entire Software Lifecycle –** To enable sustainability after a project, the Institute should work with developers during a project. Example activities could include promoting best practices, providing test harnesses, standardizing documentation guidelines, and delivering software engineering education. ***The Institute should support the entire lifecycle of software in supporting production science and indicate how they will accomplish the support and transition the software throughout the software lifecycle.***

- **Support for Entire Project Lifecycle –** ***Institute founders should be required to very clearly walk an example project through their proposed institute from innovation to sunset, including metrics and education elements.*** One idea is for Institute founders to document an example project within their Institute as a trial balloon moving through the Institute over time to elicit the attributes of the mechanics and flow of the proposed Institute. Exactly how is sustainability achieved? Outline who is working on code and projects over time and differentiate between Institute staff and non-Institute members of the community. In describing their Institute, respondents should elaborate their use of popular or common approaches. For example, many workshop attendees gleaned from Gunnar Hellekson's (of Red Hat) plenary presentation that success in *open source* encompasses much more than choosing the right license and making the source code openly available under that license. ***Institute founders should state explicitly how incorporation of a proposed concept, such as open source, whether original or borrowed from industry, will lead to sustainability, innovation, and new science.***

- **Ratings –** There was much discussion about how the Institute would provide some sort of rating or assessment of aspects of software, projects, developers, and possibly even whole communities. Each rating scenario discussed included levels for the entity to graduate through as a function of capability and maturity. Various analogies were made to current assessment models. For example a *Software Readiness Level* was proposed that might be based on the *Technology Readiness Level* used by government agencies. The Carnegie Mellon University *Capability Maturity Model* was discussed to assess software project processes. To assess whole communities, the idea of a *Cyberinfrastructure Maturity Model* was discussed as loosely based on the *Open Group Service Integration Maturity Model*. The Apache Software Foundation model was discussed. In each example, the benefit was to level set entities of a similar nature in a common assessment framework to assess capability and maturity and thus inform the communities according to stated goals and objectives. ***The consensus was that the Institute***

should continually provide substantive assessment approaches of software, projects, developers, and communities and continually make this information readily available.

- **Standards –** Workshop attendees discussed that standards, while providing benefit, are currently overwhelming because of the number of extant standards, and the challenges in terms of identifying which of hundreds to use, how to use them effectively, and why. ***The Institute should provide a substantial benefit not only in terms of collaborating with existing standards-setting organizations to create new standards, but in cataloging and identifying which extant standards to use, and providing associated software promoting interoperability across communities.*** The analogy given was for the Institute to serve as a version of an "Underwriters Laboratory" for vetting standards.

- **Governance Representation –** An environmental observatories Institute should implement a governance structure that is representative of the community across multiple dimensions including, for example, observatory type, codes, nature of research processes, education, analytics, and information technology. There should be diverse, multi-sector participation with comprehensive expertise in the represented areas. ***The Institute governance should be driven by and responsive to the community that the software is intended to serve.***

- **Community Architecture –** The consensus was that an ideal Institute approach would not be an "*uber*-institute" with a one-size-fits-all technical solution, but rather an effective distributed network of shared interoperable, extensible, and modular infrastructure; policy; governance; and community and virtual organization best practices. ***The Institute should implement a community architecture which clearly states the processes and policies by which:***
  - ***user input and requests are accommodated,***
  - ***current and emerging requirements are managed,***
  - ***community direction is assimilated,***
  - ***priorities are determined,***
  - ***development and/or development direction is provided,***
  - ***specifications are established,***
  - ***codes are matured,***
  - ***certifications are granted,***
  - ***standards are established or propagated,***
  - ***governance is established and evolved,***
  - ***various committee members are appointed or elected,***
  - ***special cross-cutting projects, workgroups and task forces are instantiated,***
  - ***effective incentive and reward structures are continually refined,***
  - ***self-sustainment is achieved, and***
  - ***other processes and policies are developed and implemented as needed.***

  Workshop attendees asserted that it is important to involve end users in most if not all of these activities.

- **Software Infrastructure –** ***There was consensus that the Institute, as an institute, should preclude serving the sole role as a primary developer of major software.*** There were examples discussed during the workshop that suggested this approach falls short of the intended vision and scope of the Institute. ***The recommendation is for the Institute to preclude the building of monolithic "one-size-fits-all" software, but instead to allow the creation of software infrastructure efforts like Software as a Service (SaaS) environments in response to community requirements as appropriate. The aim is to emphasize the sustainability and interoperability of software and not the building of new, large monolithic software.***

- **Expanded Computation and Analysis –** Much environmental observatory computation and analyses are confined to desktop PCs or small workgroup servers. An ongoing significant challenge to the environmental observatories is that they often lack some or all of the required staff, knowledge, tools, time, infrastructure, or institutional support to migrate to modern HPC architectures, grid, and/or cloud environments. Correspondingly, environmental observatories are not taking full advantage of many larger more sophisticated visualization and analysis tools associated with modern large-scale distributed computing. ***An environmental observatories Institute will be integral to realizing the linking of environmental model and data products to emerging computational and analysis environments. The Institute should assist with transitioning software from desktops to modern HPC architectures, grid and cloud environments.***

- **Permanence –** Another issue that was discussed was institutional permanence, noting that environmental observatories as instantiated last longer than 10 years. Accordingly, there was a discussion on how to attract and retain graduate students to participate in the Institute for the long-term when only finite-term Institute funding may be available. An idea put forth was for the Institute to employ "programming fellows" (described also in Section 2.3) whose role would be to support various aspects of code sustainability beyond the more transient nature of grad students. ***The longer-term "NCAR" sustained funding model was discussed as a possible model for the Institute, and the workshop recommendation was that the NSF either needs to do long-term support for the Institute, or require a plan on how it will transition away from funding after a finite term of funding.***

2.3 What expertise and capabilities should the Institute provide and how should it interface and interact with science communities? What education and outreach functionalities are meaningful in the Institute?

- **An Intellectual Hub for Synthesis –** One mechanism that the workshop attendees identified for how an environmental observatories Institute should interface and interact with science communities was serving as an intellectual hub for synthesis. As an example, organizations like the National Evolutionary Synthesis Center host synthesis activities; one such activity described as "an intense event at which a group of programmers with different backgrounds and skills collaborate hands-on and face-to-face to develop working code that is of utility to the community as a whole. The mix of people...include domain experts and computer-savvy end-users."[21] During synthesis workshops, workshop attendees would promote intense, hands-on multi-disciplinary coding and/or operational sessions that endeavor to push the boundaries of broader data and resource incorporation and analysis towards identifying the elements of transformative science and research. ***The Institute should state specifically how it will enable transformative new science.***

- **Education and Outreach –** ***Workshop attendees agreed that the Institute needs to play a strong role in education and outreach as it engages with science communities to train computationally competent scientists and engineers. Education should target students, the PIs and managers of environmental observatories, and all other end-users. Outreach should extend beyond the environmental observatory communities via linking with environmental practice and other research and data provider communities, including other Federal agencies. Outreach should promote software that is modular, customizable, and extensible to enable interoperability. Software reuse and repurposing of applications and methods should be taught. Institute***

[21] National Evolutionary Synthesis Center Hackathon CFP; Retrieved October 14, 2010 from http://gmod.org/wiki/GMOD_Evo_Hackathon_Open_Call

outreach should also facilitate software adoption among communities and must have end users involved in software development from the onset.

- **Expanded Communication –** ***The Institute should facilitate expanded communication among and between the user communities, particularly including the social sciences. Communication should also be facilitated between domain and developer communities, including domain and computer science terminology and concepts.*** This could include a common vocabulary for education and outreach supporting data storage, data ingestion, data visualization, and so forth and also concepts like domain-specific languages, frameworks, and language-oriented programming. The idea of a "distinguished fellows" program was put forth where scientists in residence would facilitate inter-community communication and collaboration to ensure effective coupling and flow of information across communities.

- **Open Source Mechanics –** ***On the development side, open source should be explicitly supported and promoted.*** Models can be borrowed from successful open source examples including Apache Software Foundation and Microsoft Outercurve. ***The Institute should provide expertise to ensure long-term sustainability of the software efforts it facilitates, including perhaps some type of certification that qualifies software for interoperability and sustainability. The Institute should provide support for the creation of software foundations that build upon successful open source practices.***

- **Programming Fellows –** A concern that was expressed was how to get around the issue of graduate students leaving before software is fully developed from the Institute perspective of interoperability and sustainability. One potential solution was for the Institute to create "programming fellows" who support community codes long-term with sustainable practices including open source and good software engineering. Such software might be transitioned from "program to product" if the Institute supported the original developers of software with promise to work directly with programming fellows in residence at the Institute.

- **Student Support –** ***The Institute must support students.*** One idea suggested was to bridge gaps by looking at where students could eventually be employed upon graduating and then facilitate the positioning of students for these destinations while supporting Institute community codes. Target destinations for students could include EPA, USGS, NOAA and NASA, for example. The Institute could offer internships for students including opportunities for domain students to learn computer science, computational/cyber-science and engineering, and software engineering topics. The Institute could create computational curriculum modules and implement a train-the-trainers model.

- **Boot camps –** ***There was consensus among workshop attendees that the Institute should provide a "boot camp" to educate and train end-users in all facets of the environmental observatory cyberinfrastructure ecosystem.*** Topic areas could be domain oriented, computer science oriented, or other. For example, there is a large gap between professional developers and academics. How does the Institute educate the community to transform the scientific software development community? Boot camps would be one approach. The "boot camp" connotation is meant to evoke that it is low cost, group-oriented, intense but reasonable to traverse for a diverse audience, and of sufficient duration to address a variety of issues and achieve a broad range of objectives.

- **Computational Science Education –** A critically significant educational challenge in computational science was emphasized at the workshop. Computer scientists in the U.S. are not being adequately educated in emerging architectures such as heterogeneous many core systems.

Further, domain scientists are not being educated in good programming practices, software and data lifecycles, computational data analysis techniques such as data mining, and sustainability methods such as key aspects of software engineering best practices. Computer and domain scientists are not being adequately educated in team approaches to computational science involving computer scientists, domain scientists, and numerical analysts responsible for best algorithm selection. The problem is exacerbated by the absence of these effective and knowledgeable computationally-oriented teams to address refactoring existing scientific code or creating new scientific code so that scientific codes can be readily updated and sustained, new capabilities can be easily added, new platforms can be supported, results can be repeated, new communities of developers can be added, and elements can be reused over long periods of time. A recent article in the October 2010 issue of Nature[22] reinforces the fact that most science code - while functional and often of high value - is not written according to computer science best practices and therefore is not sustainable, reusable, and often is not repeatable. ***Resolving the computational science education problem needs to go beyond training current computer and domain scientists and reach all the way into the undergraduate and graduate classrooms. Given the enormity of the problem, the recommendation is that the Institute should begin to address this issue in order for the U.S. to maintain international computational science leadership.***

2.4 What are the critical linkages between the Institute and other components of a community cyberinfrastructure (i.e., software tools, databases, instruments, etc.)? What is the unique role of the Institute in the broader cyberinfrastructure ecosystem (e.g., XSEDE/TeraGrid/XD, DataNet, MREFC, etc.)?

- **Linkages and Coordination –** Currently there is no overall coordination among the environmental observatories. While some of the environmental observatories individually interact with other NSF efforts like XSEDE/TeraGrid, DataNet, MREFCs, and others, it follows that there is no overall coordination between environmental observatories and these projects. Additionally, there is no overall coordination between environmental observatories and other federal agencies such as EPA, USGS, NOAA, USDA, NASA, or Army Corps of Engineers, although some individual projects (e.g., CUAHSI Hydrologic Information System) have partnerships with agencies. The coordination that is desired is to promote greater software and data interoperability in this broader cyberinfrastructure ecosystem. Interoperability is a grand challenge for the environmental observatories. Subsumed in this grand challenge are the management of current standards; development of data-driven tools and infrastructure exploiting standards and high-performance computing resources; promotion and adoption of good software engineering principles; addressing all aspects of the software and data lifecycle; effective governance, community architecture, and open source models; and effective education, training, outreach, and workforce development. Given that applications in the environmental observatories read and write data, enabling new science is a direct function of the usability of this data and ergo the software that works with this data. Environmental observatories Institute critical linkages will rely on collaboration through data, and this implies recognizing the relationship between sustainability of data collections (including real-time sensor feeds) and sustainability of the software tools that work with these. ***The Institute should state clearly how they it sustain software in a highly dynamic cyber-ecosystem.***

[22] Merali, Zeeya. Computational science: ...Error ...why scientific programming does not compute. NATURE, October 14, 2010, Vol. 467, pgs. 775 - 777. Retrieved October 14, 2010 from http://www.nature.com/news/2010/101013/full/467775a.html

- **Value of Industry Partnerships –** Several vendors were represented at the workshop including Red Hat, Microsoft Research, IBM, ESRI, and DHI. It was clear to attendees that industry partners are essential to the success of the Institute and should be required. Gunnar Hellekson, Red Hat's Chief Technology Strategist, provided a 1.5-hour presentation and discussion at the workshop on Open Source Mechanics that informed attendees of many facets and dimensions of open source that attendees had not known. Microsoft Research discussed an environmental informatics initiative and also an educational initiative with project attributes that would be difficult for a community to replicate on its own and virtually impossible to replicate and sustain across communities. Another angle on industry partnerships was the idea of the Institute utilizing professional commercial software developers and/or development firms in coordinating and servicing certain community development requirements. The Institute can uniquely and critically serve the role of effectively assimilating and disseminating the value of industry partnerships across communities. ***The Institute should require industry partnerships and clearly describe how the value provided by these partnerships will be assimilated and disseminated to across communities.***

- **Education, Training, and Workforce Development –** ***As elaborated in Section 2.3, education, training, and workforce development to support and sustain internationally competitive computational science in the U.S. is an enormous challenge and a critical need for the environmental observatory communities. From the perspective of critical linkages, the recommendation is that if the Institute effort is not the NSF vehicle responsible for addressing this, then it is paramount that the Institute's link to the NSF efforts that are responsible for addressing workforce development in a very significant way.***

- **Providing Catalogs –** ***There was strong consensus among workshop attendees that the Institute serving environmental observatories should perform the role of cataloging various software and data elements. This should include discovery, a measure of code and data quality, who is using it and how (e.g. usage models and user groups), and who is sustaining it and how. The cataloging of software and data should facilitate discoverability and interoperability across communities. What is the QA process for software and data, how is it characterized, and how will these functions be sustained over time? The cataloging function does not imply that the Institute itself should necessarily host the software or data, but rather that it serves the role of cataloging what the communities may themselves host. The catalog should offer users the assurance that the cataloged elements are supported and will continue to be there over time. The Institute should also catalog people, that is, identify the location of experts on a particular topic or technology. The Institute should promote good software engineering practices and promote software for data interoperability as a function of technology, policy, and community practice, as stated earlier.*** There was an interesting idea that the NSF could use the Institute to catalog and promote discoverability of all NSF projects and programs beyond what is provided by the current NSF website. For example, a user search across *all* NSF projects and programs via the Institute could yield more tangible assets (data, software, models, publications, etc.) for the purposes of incorporating and/or interoperating with that project or program's technology as appropriate. This cataloging, discoverability, and interoperability theme of the Institute can span many areas and communities in a high-value manner not provided for by current NSF efforts.

- **Service and Resource Provider –** The Institute could serve the role as a provider of storage for catalogs, cloud computing services provider, build-and-test environment provider, code hardening services provider, provider of services to make data more robust, and so forth. ***The Institute should state clearly what services and resources it would make available to communities.***

2.5 What are meaningful metrics, evaluation mechanisms and governance structures for the Institute? What are appropriate approaches to sustainability of the Institute?

- **Adoption and Effectiveness –** ***A primary metric identified at the workshop was adoption. Here adoption could refer to adoption of codes; best practices in software engineering; open source models; outreach, education, and training models; incentives and rewards; and standards. The adoption should be within and between communities, and could also include industry adoption. Industry adoption could include the aforementioned aspects along with intellectual property considerations, commercialization, patents, and spin-offs. Adoption should be measured as a dimension of effectiveness. There was discussion that new metrics should be able to evolve with the Institute based on experience and success, and that there should be a plan for adapting metrics to a changing environment.***

- **Impact –** Tied to adoption is the assessing of impact. Achieving interoperability is a key facet of an environmental observatories Institute on multiple levels; therefore evaluating interoperability is tantamount to success. Many of the metrics discussed were geared towards critical Institute outcomes such as achieving new science. The Institute should be able to measure how it uniquely enables the solving of more complex and/or more significant research problems than possible without the Institute. ***The measurement of new science resulting from the Institute is critical, for example number of published papers as a result of having expanded cross-cutting access to the multi-community resources that the Institute enables. An important evaluation metric is to acknowledge, measure, and assess the impact of the Institute's role in enabling its partners and users to produce quality software.***

- **Evaluation –** Evaluation mechanisms should clearly measure against the metrics and ultimately distill out from multiple vantage points what is working within the Institute cyberinfrastructure. For example, ***how does the Institute evaluate that it is choosing the right projects and activities to benefit its users while achieving its sustainability and transformative new science objectives?*** Third party firms could be employed in evaluation. ***Involving end-users in the evaluation process is also critically important. With respect to governance, metrics should be designed to allow the course of the Institute to self-correct as it is measured and assessed. Education and training metrics should enable evaluation of both quantity and effectiveness.***

- **Community Ownership –** The Institute will provide many high-value critical capabilities, functions and services to communities, however it is also important for the communities to express some ownership of the Institute. How are the communities demonstrating that it is important for the Institute to continue its role and to continue to sustain? What are communities contributing back to assure the Institute's sustenance? ***The Institute should clearly describe mechanisms and metrics of community ownership of the Institute that contribute to its sustainability.***

- **Governance Structure Effectiveness –** ***The Institute should outline its governance structure and carefully explain how the governance will be effective for addressing the community needs they are serving, ensuring that the software ecosystem is requirements driven rather than technology driven. The governance structure needs to allow for models where representatives from the various environmental observatories can be included in the Institute, with technical personnel serving on the development team, as appropriate, and end users determining the overall direction and requirements that the software ecosystem will address. It should describe how governance enables the proposed software ecosystem to be extensible and responsive to the community. For environmental observatories, the community subsumes both the environmental observatories' researchers and staff and their many and varied users. The***

governance plan should indicate how end users will be included, and it should discuss how prioritization will occur when workload exceeds available resources.

- **Opportunity and Innovation –** When a center or institute is in the position of providing resources and/or services, end-users are expected to take reasonable advantage of these offerings and correspondingly advance their research to new levels of achievement. Certain researchers are able to build on early successes, reapply for more services based on early successes, and see their successes snowball if they can continue building on prior individual or cumulative successes. On the one hand, this is beneficial for the researcher and the science represented, possibly benefitting the entire community. On the other hand, a new researcher not in a position to take advantage of early access to center or institute resources and thereby build on early successes may be less able to compete for support of a new program when there are finite resources to allocate, yet may have superior potential with a novel approach. How can the Institute make certain and be able to measure that it is always giving opportunity to new higher-risk researchers who are possible engines of innovation? ***The Institute should address how it will offer new higher-risk researchers the opportunity to partake of the Institute's offerings in balance with supporting more accomplished, but lower-risk researchers. The Institute should emphasize serving a wide gamut of multidisciplinary users and not just a small niche of users and types.***

- **Approaches to Sustainability –** Taking into consideration the tremendous transformative value an environmental observatories Institute will provide its community as described herein, it is reasonable to conclude that the value will not diminish over time. Specifically, the Institute can be thought of as an engine of continual transformative new science where the cylinders are coordinated interoperability enabled by the Institute across communities, the success and efficiency of the design is based on open source mechanics, and the air/fuel mixture is a relative never-ending abundance of data combined with finite funding. The proposed software ecosystem combined with effective governance and community architecture needs to achieve sustainable fuel. The need for this engine has been commensurate with the existence of environmental observatories since their instantiation, and the absence of this engine has been the heretofore absence of continual, community cross-cutting, transformative new science. Community members, as owners and drivers, need to own and direct the sustenance of the Institute engine. The engine metaphor is intended to exemplify how all the various parts - including owners, drivers, and the NSF - must work together to achieve sustainability that drives continual transformative new science, and that is key to any approach to sustainability.

- **Expanding the Scope of Sustainability –** The Institute should state whether it will contribute to the sustainability of other NSF initiatives. For example, will the proposed Institute contribute to the sustainability of XSEDE/TeraGrid, OOI, DataNet, etc., or perhaps even go beyond the NSF to include projects or programs within USGS, USDA, Army Corps of Engineers, EPA, and other organizations.

2.6 How would the Institute impact the science and engineering in the environmental observatory community and its practices, capabilities and productivity?

There was consensus among workshop attendees that the Institute for environmental observatories would greatly benefit their science and engineering and associated practices, capabilities and productivity by creating a cross-cutting cyberinfrastructure that enables interoperability, collaboration, and knowledge sharing at multiple levels.

- **Increased Value and Usability –** A key outcome is that the value and usability of existing data and software assets would be increased and consequently science would be transformed. The

Institute for the environmental observatory communities was determined to be critical to enabling new science and engineering that could only be accomplished by a Software Institute and not via traditional large project-oriented grants.

- **Coordination –** There currently is no effort to coordinate among or integrate environmental observatories. No one environmental observatory can "own" this role as the practices and capabilities of any one are not repurposable enough to encompass the others, and new science is not being realized for this reason. An environmental observatories Institute would enhance and accelerate new science within these communities through access to, and interoperability with, other environmental observatory assets. Furthermore new science would also be generated between communities that would otherwise not be possible. New science between communities would result from the combination of interoperability of resources at multiple levels combined with coordinated collaborative activities like synthesis workshops and working groups.

- **Community Impact –** Part of understanding the impact of the Institute requires understanding who the customers are and assessing what the impact of a successful Institute would be to them. The workshop attendees used this model to assess Institute impact on producers and consumers of environmental observatory resources as customers. Examples of impact included providing: use cases and user outreach guidance; software and data that will be put in front of many users in order to increase its usability; innovation; help in generalizing software and making it more robust; help in creating software for multi-observatory services; and incentives for researchers to contribute data and software.

- **Transformation of Science –** Part of the Institute-enabled transformation of science would result from more scientific software - of higher quality and greater usability - from more community sources with more data interoperability tools and requiring fewer transformations.

- **New Levels of Achievement –** Whether individually or as a community, the Institute should enable new levels of achievement not otherwise possible. For example, a National Water Model in part requires standards and their effective use, effective cataloging of computational models and data, software engineering best practices to allow models to be discoverable in a cataloging system and to be made interoperable at the component and/or physics level, the coordination of data and analysis across many communities, and more. In effect, achievement requires most of the offerings and benefits described in this document. In the National Water Model example, both individual researchers and the community as a whole will reach new levels of achievement, and many more achievements would be made possible by the Institute.

- **Strategic Use of Standards –** There would be more strategic use of existing standards across multiple communities, providing incentives for greater industry and other Federal agency participation.

- **Sustainability through Open Source –** A more informed use of the open source model would enhance development and sustainability of efforts that transcend students, postdocs, faculty and staff of environmental observatories.

- **Improved Quality –** The Institute for the environmental observatory community will introduce software engineering best practices, assessment frameworks, refined open source models, strategic use of standards, synthesis workshops, superior analysis and interoperability tools, and more to its community. Quality will be improved across several community dimensions.

Appendix A
Workshop Participants

Stan	Ahalt	RENCI, U of North Carolina at Chapel Hill
Dan	Ames	Idaho State U
Amy	Apon	U of Arkansas
Matt	Arrott	UCSD, Calit2
Anthony	Aufdenkampe	Stroud Water Research Center
Peter	Backlund	NCAR
James	Brunt	U of New Mexico
Randy	Butler	NCSA, U of Illinois
Scott	Collins	U of New Mexico
Judy	Cushing	Evergreen State College
Charlie	Dow	Stroud Water Research Center
Dave	Gallaher	NSIDC, U of Colorado
Ken	Galluppi	RENCI, U of North Carolina at Chapel Hill
Corinna	Gries	U of Wisconsin
Gunnar	Hellekson	Red Hat
Rick	Hooper	CUAHSI
Jeff	Horsburgh	Utah State U
Ray	Idaszak	RENCI, U of North Carolina at Chapel Hill
Marcelo	Lago	DHI
Kerstin	Lehnert	Columbia U
Yong	Liu	NCSA, U of Illinois
Alex	Mahalov	Arizona State U
John	McGee	RENCI, U of North Carolina at Chapel Hill
Mike	McGuire	U of Maryland
Anna	Michalak	U of Michigan
Barbara	Minsker	U of Illinois, NCSA
Reagan	Moore	U of North Carolina at Chapel Hill, RENCI
Larry	Murdoch	Clemson U
Marian	Muste	U of Iowa
John	Orcutt	Scripps Institution of Oceanography, UCSD
Mark	Parsons	NSIDC, U of Colorado
Beth	Plale	Indiana U
Arcot	Rajasakar	U of North Carolina at Chapel Hill, RENCI
Mohan	Ramamurthy	Unidata, UCAR
Charles	Schmitt	RENCI, U of North Carolina at Chapel Hill
Scott	Sinno	NASA NCCS
Sheila	Steffenson	ESRI
Allison	Steiner	U of Michigan
Dave	Tarboton	Utah State U
Robert	Tawa	NEON
Claire	Welty	U of Maryland
Mark	Williams	U of Colorado
Yan	Xu	Microsoft
Jeff	Zais	IBM
Ilya	Zaslavsky	SDSC, UCSD